DOCUMENT RESUME

ED 094 360 CS 001 262

AUTHOR	Myers, Jerome L.; And Others
TITLE	Memory for Prose Material. Final Report.
INSTITUTION	Massachusetts Univ., Amherst.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
BUREAU NO	BR-1-0350-FR
PUB DATE	74
GRANT	OEG-1-71-0109(508)
NOTE	143p.
EDRS PRICE	MF-$0.75 HC-$6.60 PLUS POSTAGE
DESCRIPTORS	*Cognitive Processes; Critical Reading; Educational Research; *Learning Processes; *Memory; Prose; *Recall (Psychological); *Retention

ABSTRACT

The experiments described in this report attempt to further understanding of the storage and retrieval of factual information. In the first five experiments, organization of simple prose materials is varied. Effects upon both accuracy and order of recall are discussed in terms of retrieval strategies determined by organization. The fourth and fifth experiments consider transfer effects in prose learning as a function of organization. The sixth experiment considers how the context of the prose passage influences retention of sentences embedded within it. The remaining four experiments consider the retrieval of the simplest type of proposition: "Response j is paired with stimulus i." Using time to recall the response, several alternative theories of retrieval were tested. Good fits were obtained employing a mathematical model of a strength theory. (Author)

ED 094360

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BEST COPY AVAILABLE

FINAL REPORT

Memory for Prose Material

National Institute of Education Project No. 1-0350

Jerome L. Myers

University of Massachusetts

Amherst, Massachusetts 01002

1974

The research reported herein was performed pursuant to a grant contract with the National Institute of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

BEST COPY AVAILABLE

CS 001 262

Table of Contents

page

Summary..... 1

Effects of Prose Organization upon Free Recall..... 2

The Effects of Type and Sequencing of Prose Organization on Retention..... 26

The Role of Comprehension in Learning Concrete and Abstract Sentences..... 43

Retrieval Processes in Recall..... 77

List of Tables

page

Effects of Prose Organization upon Free Recall
Table 1. Dimensions of Passages..... 7

The Effects of Type and Sequencing of Prose Organization on Retention
Table 1. Topic Matrices..... 35

The Role of Comprehension in Learning Concrete and Abstract Sentences
Table 1. Analysis of Variance and Table of Means for P(Hit) Data for Groups E, C_1 & C_2..... 69
Table 2. Analysis of Variance and Table of Means for d' Data for Groups E, C_1 & C_2..... 70
Table 3. Analysis of Variance and Table of Means for P(Hit) Data for Groups E and C_1..... 71
Table 4. Analysis of Variance and Table of Means for d' Data for Groups E and C_1..... 72
Table 5. Analysis of Variance and Table of Means for P(Hit) Data for Groups C_1 and C_2..... 73
Table 6. Analysis of Variance and Table of Means for d' Data for Groups C_1 and C_2..... 74
Table 7. Analysis of Variance and Table of Means for d' Data for Groups E and C_1 Experiment II..... 75
Table 8. Analysis of Variance d' Data for Groups E and C_1 Across Experiments I and II..... 76

Retrieval Processes in Recall
Table 1. Parameter Estimates of Strength Models..... 98
Table 2. Stimulus and response Probability Effects (Experiment III).....112
Table 3. Stimulus and response Probability Effects (Experiment IV).....120

List of Figures

page

Effects of Prose Organization upon Free Recall

Figure 1. Mean output position of correctly recalled sentences as a function of ordinal position at input, Experiment I..... 11

Figure 2. Probability of correct recall over trials for Experiment I..... 13

Figure 3. Probability of correct recall over trials for Experiment II..... 17

Figure 4. Probability of correct recall over trials for Experiment III..... 20

The Effects of Type and Sequencing of Prose Organization on Retention

Figure 1. Probability of correct recall over trials in Experiment 1..... 32

Figure 2. Probability of correct recall over trials in Experiment 2..... 37

The Role of Comprehension in Learning Concrete and Abstract Sentences

Figure 1. Comparison of P(Hit) of meaning and wording changes as a function of sentence concreteness for Begg and Paivio's data and $Control_2$..... 51

Figure 2. Comparison of d' values obtained in Experiments I and II of the present study for groups E and $Control_2$..... 54

Retrieval Processes in Recall

Figure 1. Observed and predicted RT-L functions for each trial block in Experiment I..... 87

Figure 2. Observed and predicted sequential RT in Experiment I..... 89

Figure 3. Mean RT for READ and RECALL blocks as a function of list length and AS in Experiment II.....103

Figure 4. READ and RECALL lag functions in Experiment II.....104

Figure 5. Mean RT for READ and RECALL blocks as a function of list length and AS in Experiment III.....111

Figure 6. READ and RECALL lag functions in Experiment III.....115

Figure 7. Mean RT for READ and RECALL blocks as a function of list length and AS in Experiment IV.....119

Figure 8. READ and RECALL lag functions in Experiment IV.....122

Summary

The experiments described in this report attempt to further understanding of the storage and retrieval of factual information. In the first five experiments, organization of simple prose materials is varied. Effects upon both accuracy and order of recall are discussed in terms of retrieval strategies determined by organization. The fourth and fifth experiments consider transfer effects in prose learning as a function of organization. The sixth experiment considers how the context of the prose passage influences retention of sentences embedded within it. The remaining four experiments consider the retrieval of the simplest type of proposition: "Response j is paired with stimulus i". Using time to recall the response, several alternative theories of retrieval were tested. Good fits were obtained employing a mathematical model of a strength theory.

Effects of Prose Organization upon Free Recall[1]

Jerome L. Myers, Kathy Pezdek, and Douglas Coulson

Three experiments investigated effects of prose organization on free recall. Passages consisted of five paragraphs of five sentences each and described five attributes of each of five fictitious countries. Passages were organized either by name, by attribute, or randomly. Experiment I demonstrated significantly better recall under Attribute than under Name or Random organization. Experiment II shows that only in the Attribute condition is serial recall as good as free recall, suggesting that Attribute Ss normally use serial order cues as a basis for retrieval. Experiment III verifies this hypothesis; the superiority of the Attribute condition in free recall is cancelled when serial order cues are unavailable.

The importance of organization for recall has been repeatedly demonstrated (e.g., Cohen, 1966; Cofer, 1965; Tulving and Pearlstone, 1966). However, the majority of these studies have utilized only words or word-lists as stimuli and have been concerned only with the comparison of blocked with randomized lists. Furthermore, the blocking has been in terms of categories that were presumably well established prior to S's introduction into the experiment - e.g., minerals, animals, flowers. The present research extends the list-recall work in several ways. First, simple prose materials are presented to Ss and the information to be remembered falls into categories which must be learned in the experiment. Second, the question is raised as to whether different types of nonrandom organizations of the materials are differentially effective and, if so why.

In the experiments to be reported, subjects were presented with a series of sentences, each relating a concept name (N) to an attribute (A) of that concept; for example, "The climate of Melin is humid". Melin is the name, climate is the attribute, and humid is the attribute value. Such sentences can be sequenced so that each paragraph is a set of sentences covering all the attributes of a single name (N organization). Alternatively, a single paragraph might describe the climates of five different countries (A organization). Finally, the complete set of sentences might be randomly sequenced and divided into paragraphs (R organization). Comparing these three organizations, Frase (1969) found that in free-recall tests there was little difference be-

tween A and N organized material (although the A organized material was remembered slightly better) and both led to significantly superior recall than did R organized material. Measures of clustering indicated more organization of output with N organized input than with A or R organized input, and that the A and R groups exhibited more N-clustering in output than had been present in the passages they had read.

In a second study employing N, A, and R organization of prose materials, Schultz and DiVesta (1970) found N recall better on the first of three trials, and A recall slightly better on the remaining two trials. With respect to organization of recall, they concluded that there was a pronounced tendency to organize by Name.

Both sets of authors attributed the inferior recall of R Ss to a loss of study time incurred while attempting to impose an organization upon the R materials. Frase explained the higher organization of N Ss by arguing that they could represent each paragraph by a single name and a list of attribute values; A Ss had to represent each paragraph by a list of paired-associates, k pairs of names and attribute-values. Schultz & DiVesta also appeared to believe that N organization of output is more efficient than A organization and, therefore, preferred by subjects. But if N organization of output imposes less of a memory load, as the above explanation clearly implies, why do these investigators fail to demonstrate a clear superiority in recall of N Ss over A Ss? Even if we do not accept Frase's explanation of the higher clustering in his N group, the experimental literature on free recall of word-

lists would strongly suggest that more organization implies better recall. The N groups should, therefore, recall better. This is the basic issue that stimulated us to examine more closely the designs of the two studies mentioned above, and to carry out our own comparison of N, A, and R organization of prose materials.

Experiment I

In both the Frase and Schultz & DiVesta studies, the same attribute value appeared with several names. Such materials may encourage a form of A organization which is highly efficient but not really consistent with Frase's characterization of A organization as consisting of k paired associates. Therefore, we decided that no attribute values would be repeated in our materials. Both sets of investigators used attribute values which varied in length and presumably difficulty; e.g., Schultz & DiVesta, describing type of society, used urban with one name and urban and manufacturing with another. In the present study, one word attribute values were used. Frase used fewer names than attributes; thus N Ss read fewer but longer paragraphs than A Ss. Since such distribution of material may be important and confound the basic contrast of N and A organization of input, equal numbers of names and attributes were used in the present study. Time per paragraph, rather than total time for the passage (as in the previous studies) was controlled in order to ensure equal exposure to all paragraphs; Frase's finding of better recall for earlier paragraphs might possibly be due to Ss spending more study time on the earlier para-

graphs. Finally, the order of presentation of paragraphs within each type of input orgarization was counterbalanced, an elementary precaution which appears to be lacking in the previous studies.

Method

Subjects. 45 students from an undergraduate Educational Psychology class at the University of Massachusetts volunteered as Ss. They were run in groups ranging in size from 2 to 8 Ss. Prior to testing, Ss were informed only that they were to participate in a prose-learning experiment.

Materials. The reading material consisted of 5 paragraphs of 5 sentences each. Each paragraph was typed on a separate page of a 5-page reading booklet. A separate lined booklet was used for testing. The sentences were generated by combinations of the names of 5 fictional countries (N) and 5 attributes (A) corresponding to each country. This arrangement, as specified in Table 1, is simply a 5-by-5 matrix with the names as the row marginal entries and

---

Insert Table 1 About Here

---

the attributes heading each column. The 25 attribute values (AV) were then arranged within the matrix. There were no repetitions of AV's. The country names were all pronounceable, 5 letter words, each beginning with a different consonant. All sentences were of the form: "The (A) of (N) is (AV)." This structure was utilized in all arrangements of all sentences to eliminate variability in syntax.

Design. The experimental variable was organization. The sen-

Table 1

Dimensions of Passages

Attributes / Names	Agricultural Product	Climate	Language	Industrial Product	Geography
Melin	Corn	Humid	French	Machinery	Plateaued
Pemol	Rice	Hot	German	Automobiles	Hilly
Tupel	Potatoes	Dry	English	Appliances	Coastal
Jamba	Wheat	Cold	Danish	Transformers	Mountainous
Sayon	Tobacco	Rainy	Spanish	Airplanes	Flat

tences were arranged in three organization groups -- (1) Name, (2) Attribute, (3) Random. The same 25 sentences were used in each group; only the order of these sentences varied. In the N group, each paragraph dealt with only one country, describing the 5 attributes associated with that country. The attributes in these paragraphs were in the same order on each page. For the A organization group, each paragraph dealt with only one attribute (i.e., Language, Geography, etc.). The names in the A paragraphs appeared in the same order on each page as did the attributes in the N paragraphs. In the R organization each paragraph consisted of 5 sentences chosen at random from the 25 available sentences. The order of the 5 paragraphs within all booklets was Latin Squared over subjects for all groups. Over trials, a given S used the same booklet and thus read the same paragraphs in the same order.

Procedure. Ss read the materials in timed intervals of 40 seconds per page. Thus, time was equally distributed over paragraphs. Prior to each reading, the Ss were read the following instructions.

> When you are told to do so, you will read about the characteristics of different countries from the booklet before you. You will be allowed a sufficient duration of 40 seconds to read each page. You need not hurry through the reading as this duration is more than adequate. Do not turn any page until you are told to do so. When you have completed reading the 5 paragraphs, I will ask you to write down what you have learned. This sequence will be repeated 2 times.

Immediately following each reading through the 5 paragraphs the subjects were tested for free-recall of the material. They were

instructed as follows:

> Turn over the reading booklet. You now have a sufficient duration of 7 minutes to write down everything you can recall from the entire passage. You need not use complete sentences.

Although Ss were told that they were limited to 7 minutes for the free-recall test, no group was stopped until all writing had stopped. This period never exceeded 8 minutes. Most Ss stopped within 7 minutes. This reading and free-recall sequence was repeated twice in succession. Note-taking was not allowed.

Results and Discussion

Organization of recall: cluster indices. A and N cluster indices (ARC score; Roenker, Thompson, & Brown, 1971) computed for each trial for each group, are tabled in the Appendix. As in the Frase (1969) and Schultz & DiVesta (1970) studies, the results imply better organization of output following N input than following A or R input and a preference to organize output around names rather than attributes. Unfortunately, despite consistency of results from three laboratories, the written organization of output may reflect a procedural artifact. When output is written on a single sheet of paper, the written order of output may not reflect retrieval order. In particular, in the A condition, Ss may retrieve by attribute but place all sentences sharing a common name on the portion of the output sheet, possibly because of preexperimental habits and possibly because they could then use the convenience of ditto marks instead of rewriting each name. Such ditto marks were observed both in the Schultz & DiVesta study and in the present one. Perlmutter and Royer (personal communication), using

our prose materials, had Ss write each name-attribute value pairing on a different piece of paper. They found that A Ss organized recall around attributes to about the same degree that N Ss organized recall around name. Thus, it does not appear that N input yields better organized output than A input. However, Royer & Perlmutter do confirm the previous finding that R Ss prefer to organize their output around names; their N cluster index was .40 and the A cluster index was .03. This implied preference for N organization is supported by responses to questionnaires in our experiment.

Organization of recall: order of output. Figure 1 presents for trial 3, the average position in which a correctly recalled sentence was outputted as a function of its serial position at input. Each curve represents a different input paragraph and each

---

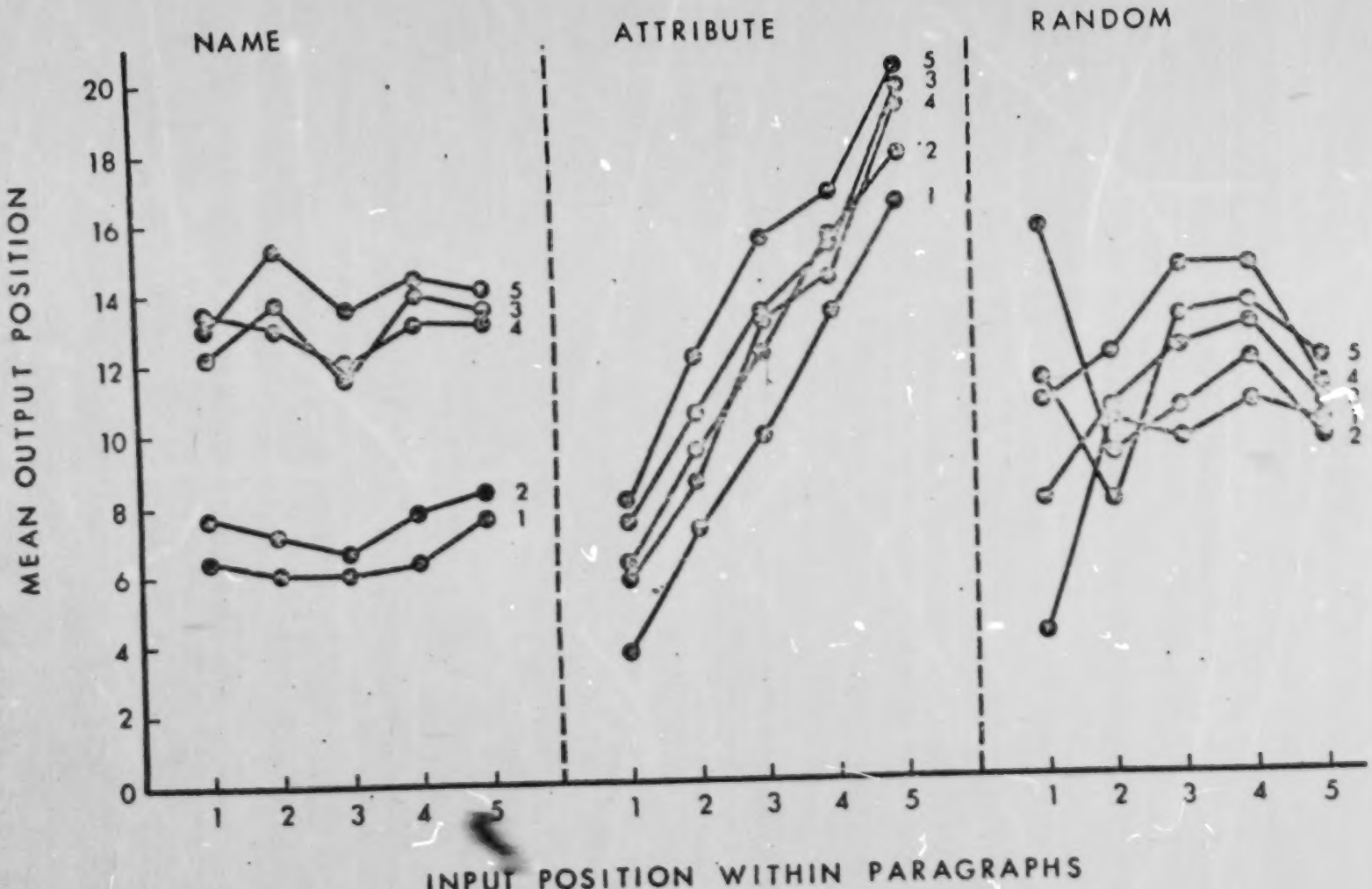
Insert Figure 1 About Here

---

point on the curve has as its abscissa coordinate the input position within the paragraph. The slopes of the curves are the important features. If the output curve for a paragraph is flat, the order of recalling sentences from a paragraph is independent of the input order. A slope of +1.0 would indicate that the output order correlated perfectly with the input order for that paragraph.

Considering N data first, the flatness of the individual curves is noteworthy. Apparently, the attribute values associated with a particular name are not recalled in the order in which they appeared in the inputted paragraph; instead, they seem to have been retrieved

Mean output position of correctly recalled sentences as a function of ordinal position at input, Experiment I



almost at random from the set of attribute values associated with a given name at input. The R output data, while considerably less orderly, tend to share this characteristic of essentially flat output curves.

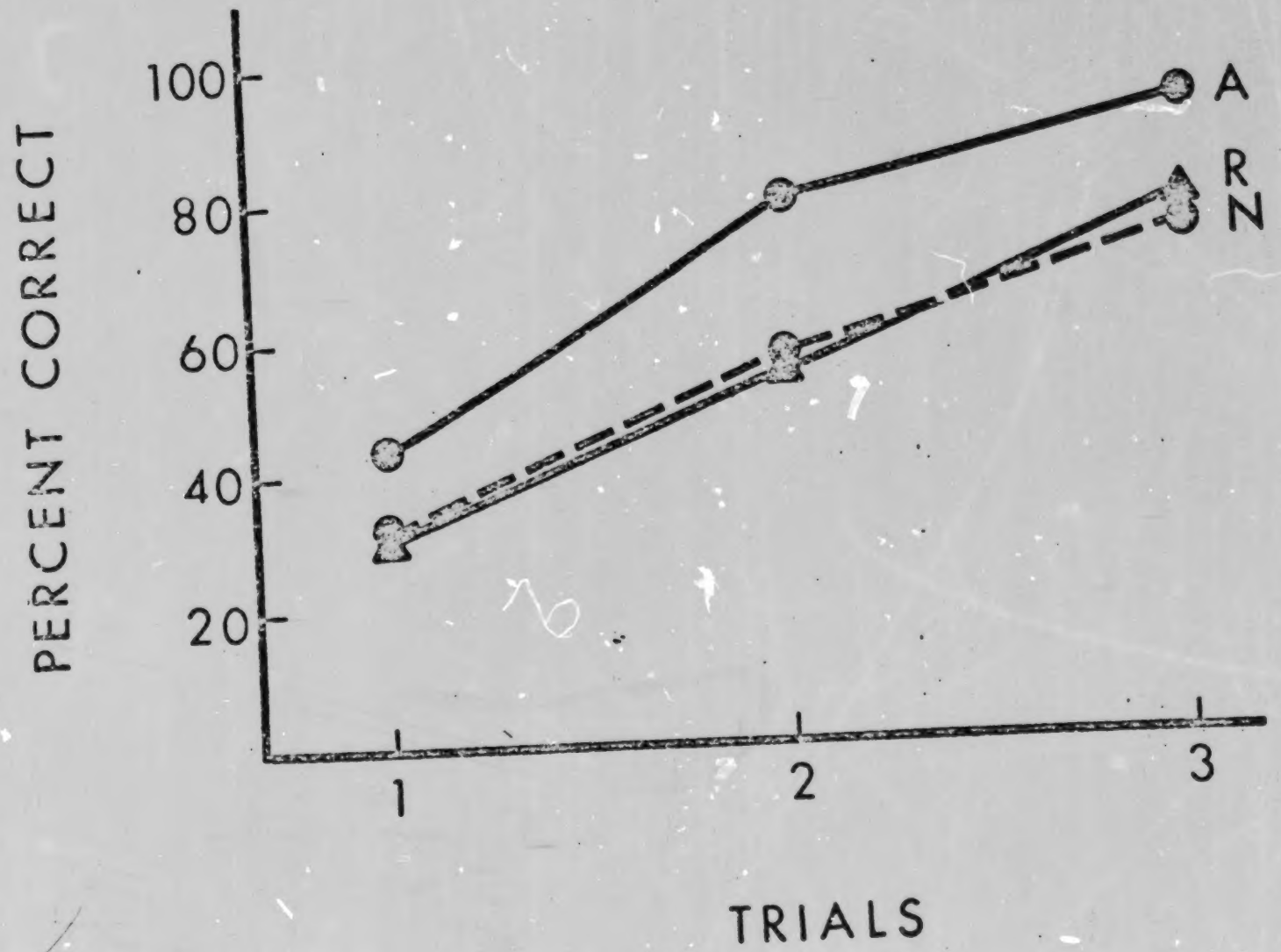
On the other hand, the A condition yields curves with positive slopes. For any given attribute, the name-attribute value pairs are recalled in the order in which the names have been inputted. Thus, serial information in the original prose is preserved in output only in the A condition.

Recall accuracy. A sentence was considered correct if an attribute value was paired with the corresponding name. Mean accuracy scored for groups, A, N, and R were 18.29 (73.2%), 13.64 (54.6%) and 13.75 (55.0%) respectively. Organization did have a significant effect on recall; $F(2,44) = 8.653$, $p < .001$. The interaction of input order and trial number did not prove to be significant. This effect can be seen in the almost parallel plots in Figure 2. Varying the order of paragraphs within each particular type of organizaional input did not affect accuracy of recall.

Insert Figure 2 About Here

The results are not consistent with those obtained by Frase (1969) or Schultz & DiVesta (1970) in earlier studies of organization of prose. They found very slightly better recall under A organization compared with N and both of these conditions much superior to the R condition. The relatively good performance of the R group in the present study may be due to a light information

Probability of correct recall over trials for Experiment I



processing load - fewer sentences than in the earlier studies, constant syntax over sentences which probably facilitates skimming in reading, and considerable time (8 seconds) to read each sentence.

At this stage in the research, the marked superiority of A over N recall was surprising. One hypothesis for this result was based on the premise that our cluster indices were valid representations of organization in storage or at retrieval (we were not cognizant of the possible artifactual basis for the marked N organization until the set of experiments were over). If so, A and R Ss, attempting to achieve N-organization output, reorganize input more than N Ss do. Such reorganization may involve additional processing of the material being reorganized, and consequently better recall than would occur without such reorganization. The next experiment tests this hypothesis.

Experiment II

This experiment was designed in response to the apparent tendency of A Ss to reorganize input. It was hypothesized that such reorganization might involve more processing of material than occurred in the N condition and consequently better recall. To minimize such reorganization, half of the Ss were run under a serial recall (SR) procedure in which they were required to output in the order in which the sentences had been input. They were informed of this before reading the material. As will be seen, the results of the SR procedure were illuminating but not

in the manner anticipated. The focus in this as in the following experiment will be on accuracy of recall; cluster indices, order of output data, types of errors, and serial position data are available upon request.

Method

Six groups of 20 Ss were run in a 3 x 2 design with input organization (N, A, R) and method of recall (FR, SR) as the variables. Ss were instructed as to their method of recall prior to the first reading and reminded before each recall and reading, thereafter. In all other respects, the materials and procedure followed that of Experiment I.

Results and Discussion

Figure 3 presents accuracy of recall for each combination of input and type of recall. The FR data are more similar to those obtained by Frase (1969) and Schultz & DiVesta (1970) than to those of Experiment I. The effect of input organization is significant, $F(2,42) = 12.76$, $p < .001$; the effect is clearly due to the superiority of the A and N conditions over the R condition, whereas the input effect in Experiment I was due to superiority of A over N and R. Despite this discrepancy, there seems little question that A organization is superior. In all studies, our own and others', there is always at least a slight superiority on trials 2 and 3. Furthermore, part of Experiment III replicated the clear difference between A and N input noted in Experiment I. Finally, Perlmutter and Royer (personnel communication), in the one set of conditions which closely replicated our N, A, and R

conditions, obtained results which are almost identical to those of Experiment I.[2]

SR data of Figure 3 are based on a lenient scoring system in which any item correctly recalled is scored as correct; a more conservative scoring system in which items must be recalled in the

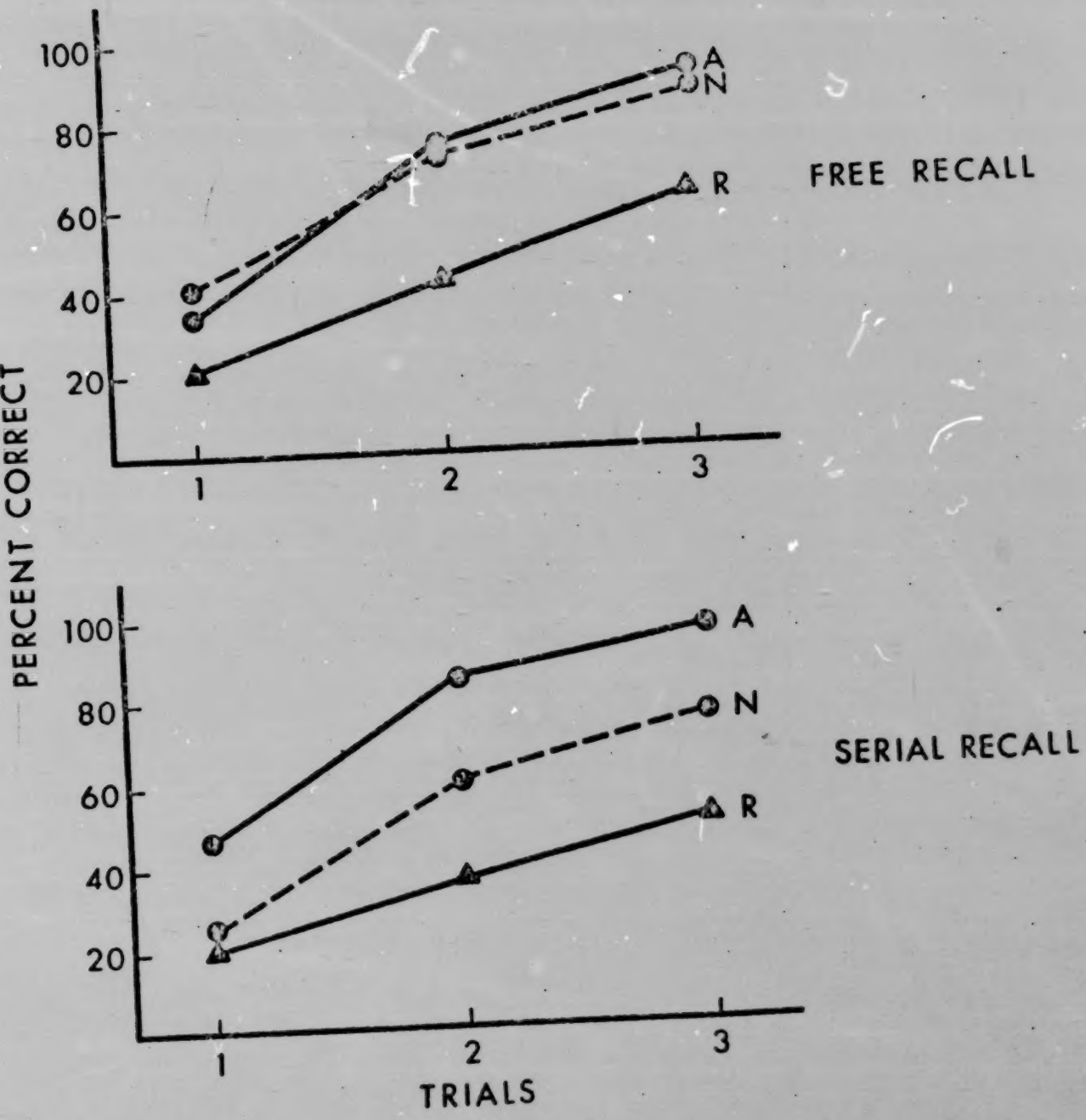
---

Insert Figure 3 About Here

---

proper order to be scored correct yields similar trends with a marked decrease in N and R scores. Under lenient scoring, A recall is, if anything, slightly better under SR than under the FR condition, N and R are considerably lower under SR. The input variable is significant, $F(2,42) = 19.57$, $p < .001$; the three input conditions are significantly separated from each other.

A review of questionnaires administered in both experiments suggests that N and A Ss are approaching their task in very different ways. N Ss describe various mnemonics including the occasional use of imagery. A Ss frequently mention that they have learned six serial lists -- a list of names, a list of climates, a list of agricultural products, and so on. This strategy is feasible because the names appear in the same order in every A paragraph. The hypothesis that A Ss develop a serial list learning strategy while N and A Ss do not is consistent with the observation that order of output preserves input order information only in the A group (see Figure 1 and the Appendix). Furthermore, if true, it is not surprising that A Ss do as well, perhaps better, under SR than under FR; the serial list strategy which is appar-



Probability of correct recall over trials for Experiment II

ently a natural one under the A condition is highly efficient for SR. The natural strategy for N Ss does not preserve serial information and SR places demands on N Ss that their approach to learning the materials does not equip them to meet. Experiment III provides a direct test of this differential strategy hypothesis.

Experiment III

If Ss in the A condition are learning six serial lists, the strategy will be negated by varying the order of names across the attribute paragraphs. On the other hand, there is no reason to expect that varying the order of attributes across name paragraphs should impair recall since neither output order nor serial recall data for this condition suggest any dependence upon serial order within paragraphs. The present experiment tested this prediction.

Method

Four groups of 20 Ss, volunteers from Psychology courses, were run. The design was a 2 x 2, with either A or N organization and with order-within-paragraphs either constant or variable. Constant order, the approach of the first two experiments, implies that attributes are listed in the same order in all name paragraphs and names are listed in the same order on all attribute paragraphs; in the variable order conditions, the order of attributes differed randomly over the five name paragraphs and the order of names differed randomly over the five attribute paragraphs. The general procedure and materials were otherwise the same as in the previous experiments.

Results and Discussion

The hypothesis that serial strategies support recall in the A condition and not in the N condition is clearly supported by the recall data of Figure 4. Organization and order variables interact significantly, $F(1,76) = 13.14$, $p < .001$. The source of the

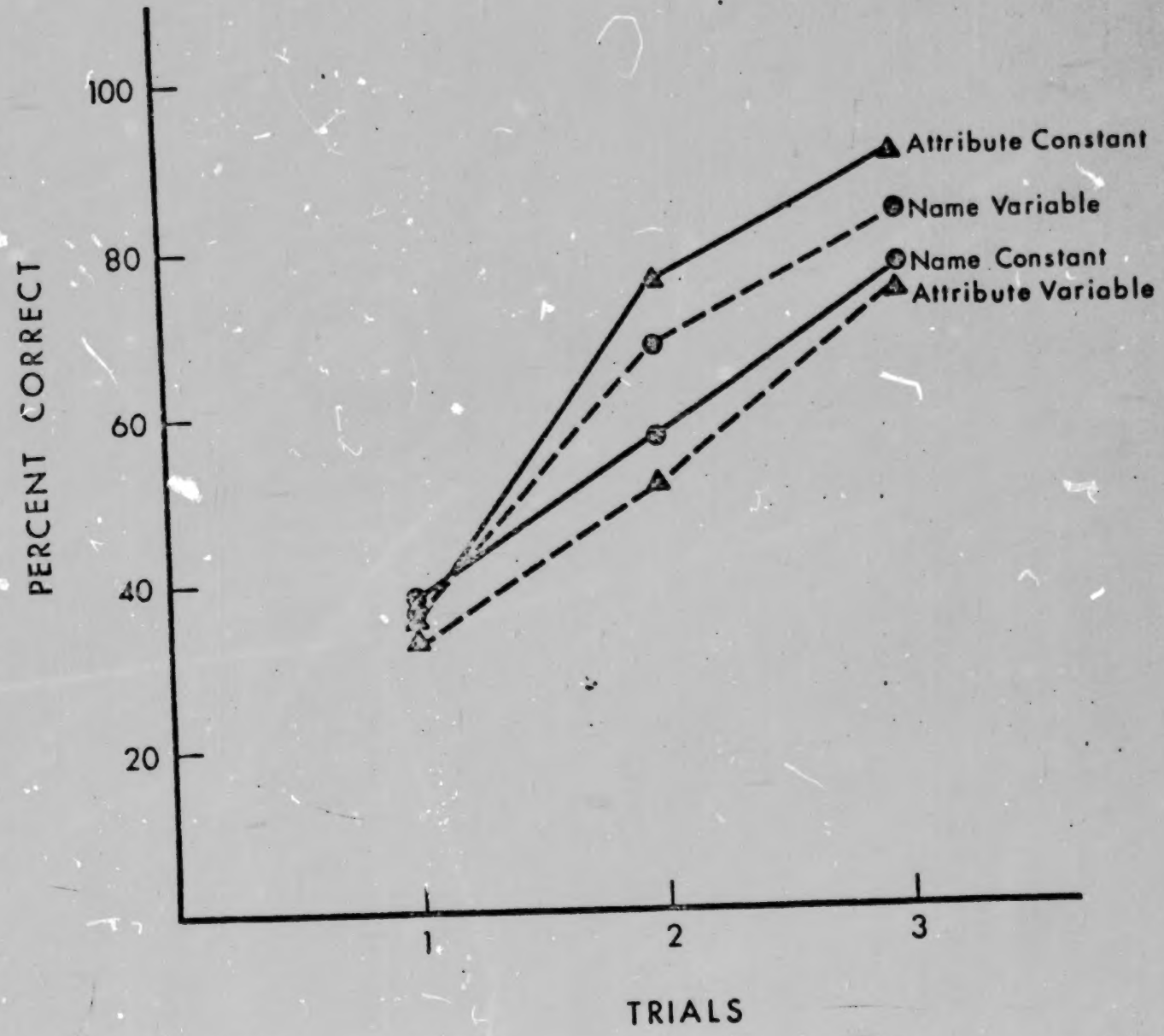
Insert Figure 4 About Here

interaction is clearly the depressed A performance under variable order; the difference, averaged over trials, is about 15%. There is actually a slight, but nonsignificant, improvement (about 5%) for N organization when order is varied. The interaction of organization x order x trials is significant, $F(2,152) = 91.73$, $p < .001$; this is presumably due to the fact that the organization x order effects do not appear until trial 2. Finally, in trials 2 and 3, under the constant input conditions, the superior recall of A Ss over N Ss observed in Experiment I are clearly replicated.

General Discussion

It is clear from the results of these three experiments that N and A subjects employ different learning strategies. N Ss have learned five unordered lists, each consisting of a name and five semantically unrelated attribute values; these Ss occasionally verbalize attempts to construct mnemonics to relate the elements of each list. A Ss have learned six ordered and semantically characterized lists - names, languages, climates, and so on. The evidence lies in (a) Ss response to questionnaires asking about

Probability of correct recall over trials for Experiment III



learning strategy, (b) the finding that order of output reflects input order only for A Ss, (c) the finding that only recall of A Ss is not impaired by the demand that information be recalled in the original input order, and (d) the finding that only the recall of A Ss is impaired by the variable order condition of Experiment III. Reexamination of Table 1 suggests why a serial list-learning strategy would be employed by A Ss and not by N Ss. If the input items in each column of the table are learned serially, the correct name-attribute value pairing can immediately be retrieved. On the other hand, learning the entries in the rows of Table 1 in serial order gains Ss nothing because the unordered set of one name and five attribute values provides sufficient information to retrieve name-attribute value pairs.

These arguments explain why the two groups use different strategies but not why the serial list-learning strategy of A Ss result in superior recall for that group of Ss. One possibility is that A Ss recall more, not because of the efficiency of the serial list-learning strategy, but because of some other factor. There are at least two such factors present in A organization and absent in N organization which, on the basis of the verbal learning literature, might be supposed to aid the A group. Under A organization, there is distributed practice on the CVCVC names, thus presumably aiding learning of those names. Also, under A organization, the attribute values are blocked by semantic category (climate, language, etc.), a condition known to facilitate recall in list-learning studies. While these two factors - distributed practice on names

and semantic blocking on attribute values - may be important in the superior recall of A Ss, they are not sufficient to account for the difference between groups. Both factors are present in the varied - A organization condition of Experiment III, a condition which resulted in rather poor recall. It appears that the superior recall of A Ss is primarily due to the presence of a serial-order retrieval scheme which is natural and easy to use, and quite effective. The problem now is one of defining the conditions under which serial order is best learned.

Certainly, there are conditions under which A organization is not significantly better than N organization, possibly because learning serial order is difficult under those conditions. Both Frase and Schultz & DiVesta obtained only slightly better recall under A organization. Unfortunately, the present study differs from theirs in so many ways, that it is difficult to specify which factors are most pertinent. In general, both of the earlier studies imposed greater information processing loads upon Ss. In particular, it may be that more and longer serial lists make the serial list-learning strategy more difficult to apply; the previous studies employed 6 x 8 and 6 x 6 name-attribute matrices compared to our 5 x 5 matrix. Another source of discrepancy may be the fact that in the previous studies several names frequently shared the same attribute value. The serial list-learning strategy employed by our A Ss may be less apparent when some set of attribute values can be encoded in a way that deemphasizes the list aspect - e.g., "all the countries increased in population except _____." In

any event, systematic investigation of procedural differences between our studies and their predecessors should provide further understanding of the role of input organization.

There may also be conditions under which A organization results in better recall for reasons other than those present in our experiments. Friedman and Greitzer (1972) used hierarchically structured material and obtained better recall under A than under N organization. There is no indication that the retrieval strategy of their A Ss involved learning serial order of either names or attribute values. Sentences were presented in the variable order which Experiment III has demonstrated provides poor recall with our materials.

References

Cofer, C. N. On some factors in the organizational characteristics of free recall. American Psychologist, 1965, 20, 261-272.

Cohen, B. H. Some-or-none characteristics of coding behavior. Journal of Verbal Learning and Verbal Behavior, 1966, 182-187.

Frase, L. T. Paragraph organization of written materials: The influence of conceptual clustering upon the level of organization of recall. Journal of Educational Psychology, 1969, 60, 394-401.

Friedman, M. P., and Greitzer, F. L. Organization and study time in learning from reading. Journal of Educational Psychology, in press.

Roenker, D. L., Thompson, C. P., and Brown, S. C. Comparison of measures for the estimation of clustering in free recall. Psychological Bulletin, 1971, 76, 45-48.

Schultz, C. B., and DiVesta, F. J. Penn. State Techinical Report, The effects of passage organization and note-taking on the selection of clustering strategies and on recall of textual materials, 1970.

Tulving, E., and Pearlstone, Z. Availability versus accessibility of information in memory for words. Journal of Verbal Learning and Verbal Behavior, 1966, 5, 281-291.

Footnotes

1. This research was supported by an Office of Education grant, OEG-1-71-0109(508). The authors wish to thank Joyce Reilly and Peter Piusz for running the subjects in Experiment II, Charles Clifton for commenting on an earlier version of the manuscript, and M. James Royer and Jane Perlmutter for several helpful discussions of the data.

2. Recently, several experiments on our laboratory have obtained better recall under A than under N organization, using different materials including common names. The superiority of A over N organization has also been obtained with different materials and 4th graders in a recent honors thesis carried out at Mount Holyoke College (Shapino, Susan: Effect of organization on the recall of written materials in fourth-grade children, 1971, unpublished).

The Effects of Type and Sequencing of Prose Organization on Retention[1]

Douglas Coulson and Jerome L. Myers[2]

The effects of several types of organization on prose retention were examined in two experiments. Of particular interest were transfer of strategy effects when each of several different sets of materials is presented once and then tested for recall. Each set of materials consisted of five five-sentence paragraphs describing five attributes of five names. Passages in Experiment 1 were organized by name, by attribute, or randomly. Experiment 1 demonstrated significantly better recall under Attribute than under Name or Random organization as well as significant improvement over trials. However, it was not clear whether the effect was due to warm-up or transfer over trials. Experiment 2, designed to distinguish between these two possible interpretations, investigated the differential effect of constant organization vs changing organization over trials. Again, a different set of materials was presented and tested on each trial. Experiment 2 demonstrated significantly better recall over trials for the Attribute Constant group; transfer of strategy is optimal under constant attribute organization.

The organization of information presented in educational materials is a major consideration in their design. The two experiments reported in this paper examine the effects of several types of organizations on prose retention. Of particular interest are transfer effects under different organizations when several sets of materials are to be learned.

In two recent experiments (Myers, Pezdek, & Coulson, 1973; Friedman & Greitzer, 1973), attribute-organized passages were recalled better than name-organized passages; in attribute-organized passages, each paragraph presents values of one attribute (e.g., "climate") for each of several names (e.g., countries) and, in name-organized passages, each paragraph presents values of several attributes for a single name. Myers et al. (1973) found (a) that the attribute-organized condition elicited an output order highly correlated with input order even under free recall instructions, (b) that serial recall was equal to free recall in the attribute-organized condition but not in the name-organized condition, and (c) that memory of attribute-organized passages suffered when the order of names varied from paragraph to paragraph. They concluded that the use of serial order cues in storage and retrieval was a natural and powerful device for Ss in the attribute-organized

condition, while Ss in the name-organized condition apparently represented each paragraph as an unordered list consisting of a name and several semantically unrelated attribute values.

In the studies cited above, Ss read the same materials on each recall trial. In the experiments reported in this paper, S was required to learn a new set of materials on each of several trials. This enables us, first, to examine the generality over materials of the finding that recall is better for attribute than for name-organized materials, and second, to investigate transfer of strategy. We were concerned with (1) whether recall improves over trials when different material is read and recalled on each trial, (2) whether such transfer, if obtained, requires a constant organization over trials, (3) whether degree of transfer is a function of type of organization.

Experiment 1

In this experiment, Ss read and attempted to recall a different set of materials on each of five trials. Organization was constant over trials. Following the argument that such passages evoke definite retrieval strategies, and assuming that Ss improve in their use of such strategies, we expected our Ss to exhibit improvement over trials even though the material was changing. Under attribute organization, Ss have a powerful strategy available to them; as Myers et al. (1973) demonstrated, they can, and do, store and retrieve the information in each paragraph as an ordered list of names and an ordered list of the attribute values which are associated with

those names. We expected attribute organization to produce the most pronounced improvement over trials as this strategy is acquired by increasing numbers of Ss over trials.

Method

Subjects. 45 students from an introductory undergraduate psychology course at the University of Massachusetts, volunteered as Ss. Approximately 5 Ss were run in each experimental session, although the range was from 3 to 12.

Materials. Each subject received a packet of 5 reading booklets and 5 response booklets. Each reading booklet contained five, five-sentence paragraphs on one of five topics. The five names used within each topic were chosen to generate a familiar but ficticious groups of sentences. For example, in the passage about brands of beans, the beans were named: Vega, Impala, Gremlin, Torino, and Maverick. All sentences were of the form: "The (A) of (N) is (AV)." An example of a prototypical sentence was "The color of the Impala bean is red". Response booklets consisted of 25 blank pages so that Ss would write only one sentence per page during the free recall response period.

Design. Three organizational groups of 15 subjects were run. For the attribute-organized (A) group, each of the five 25-sentences reading booklets was divided into five paragraphs. Each paragraph in a booklet dealt with a single attribute. For example, a single paragraph might relate where each of five types of beans were grown; "where grown" is the attribute and

its values are "canyons, hills, plateaus, valleys, and deserts". The next paragraph would consider a different attribute of these names (brands of beans). The names appeared in the same order in all five A paragraphs.

The name-organized group received reading booklets that dealt with all the attributes of a given name before introducing the next name. For example, the Vega's beans attribute (and the 5 associated values) were listed, then the Gremlin bean's, and so on. For the random organized groups, the 25 sentences for each of the 5 topics were randomly ordered and put into 5 paragraphs of 5 sentences each. Within each of the 3 groups order of topic presentation was Latin squared.

Procedure. Subjects were told that they would read five different sets of meaningful prose and be asked to write down the specific content of each set immediately following each reading. Subjects read the materials in timed intervals of 25 seconds per page. Each page contained 5 sentences of one paragraph. After each of the five trials, Ss were given 7 minutes to recall the material just read. Responses were recorded in the 25 page response booklets with one sentence permitted per page. The five trials were completed in one hour for all sessions.

Results

Free Recall. As can be seen in Figure 1, the attribute and name conditions exhibited increasing recall over the five trials, with attribute recall superior to name recall on all

trials. The random condition showed the poorest recall on all trials, first increasing to trial three then decreasing. The

---

Insert Figure 1 About Here

---

effects of organization and trials was significant, $F(2,30) = 8.42$, $p < .005$, and $F(4,120) = 17.49$, $p < .001$. There was also a topics main effect, $F(4,120) = 4.71$, $p < .005$. There were no significant interactions.

Discussion

The significant organization main effect is consistent with the previous finding by Myers, Pezdek, and Coulson (1973) that attribute organization produces superior recall than does name or random organization. Further, the fact that this result occurred with several sets of material extends the generality of the previous finding.

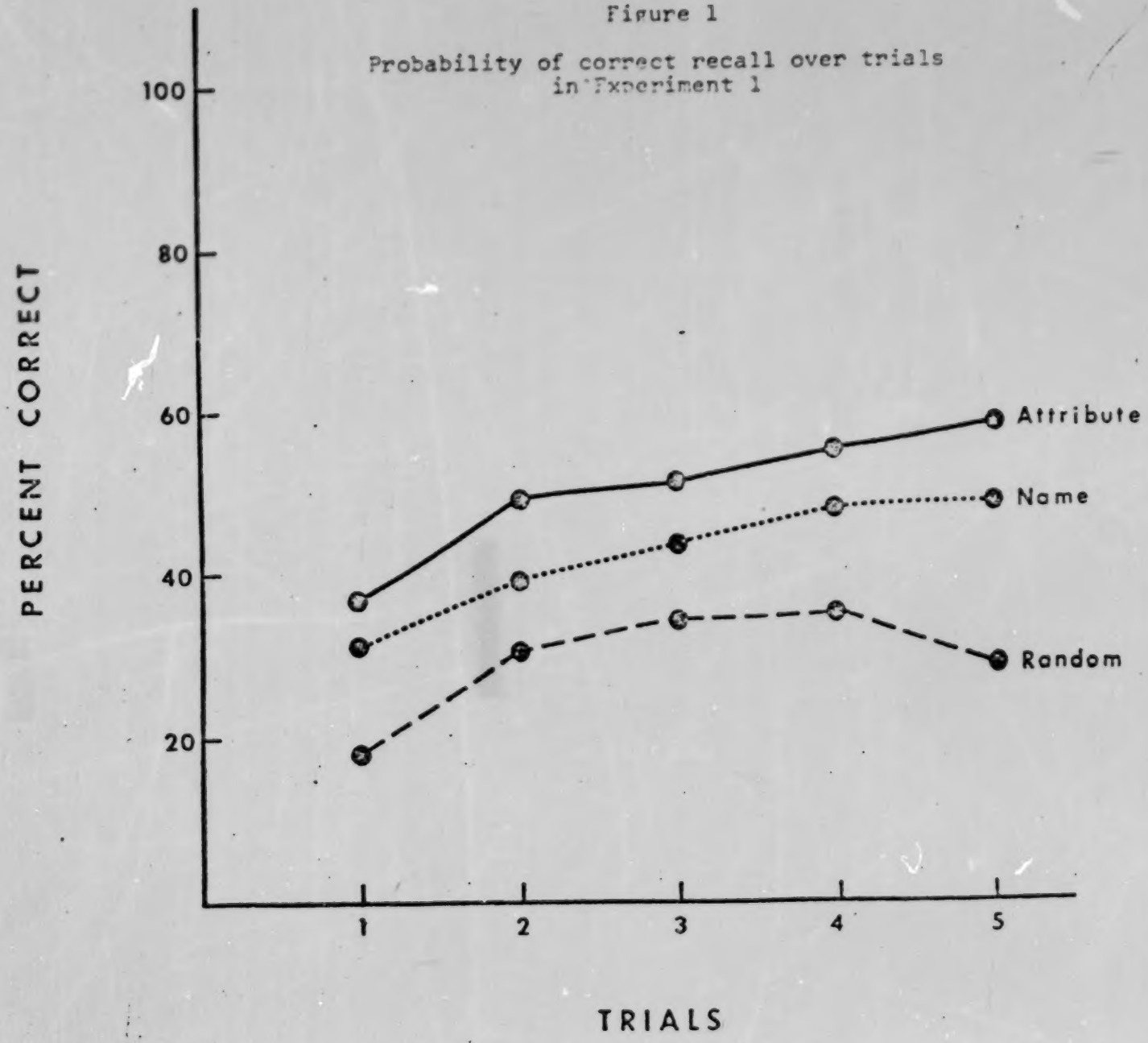
Secondly, consistent with the hypothesis that transfer occurs, there is a significant trials effect. However, the absence of a significant organization-by-trials interaction permits two possible interpretations. Either all groups improved equally in their use of a retrieval strategy or a warm-up effect was operating. Therefore, Experiment 2 was designed to distinguish between these interpretations.

Experiment 2

Experiment 2 investigated the differential effect of constant organization vs. changing organization over three trials. Both constant and changing organizational groups will be ef-

Figure 1

Probability of correct recall over trials in Experiment 1



fected equally by warm-up over the three trials. However, if a transfer of strategy is occurring, those groups that are given the same organization from trial to trial should show more improvement over trials than those groups that see a different organization from trial to trial.

In addition to introducing varied organization, several modifications in the experimental task were made to permit transfer of retrieval strategy to exhibit its effects, if it is a factor. We used CVC names to shift memory reliance to the organization itself, cut down reading time from 25 to 20 seconds per page, and reduced possible effects of rehearsal by interpolating a backwards-counting task between each reading trial, and the free recall period. All of these modifications should make learning more difficult, thus placing greater emphasis on efficient retrieval strategies.

Method

Subjects. 52 students from an introductory undergraduate psychology course at the University of Massachusetts, volunteered as Ss. Each experimental session contained approximately six Ss, although the sessions ranged in size from three to 15.

Materials. Each Ss received a packet of three reading booklets and three blank response booklets. Each reading booklet contained five five-sentence paragraphs on one of the 3 topics. The 3 topics were: Brands of Beans, Neighborhood Dogs, and Types of Candy; the complete materials are specified in matrix form in Table 1. The names used were CVC's, each start-

ing with a different consonant and each having a meaningfulness index of approximately .5 (Archer, 1960). The response booklets consisted of 25 blank pages so that Ss could write each recalled sentence on a separate page.

Design and Procedure. Ss were randomly placed in one of the four possible groups. Each group received a different sequence of organizations over the three trials. Specifically, one group received attribute organization on all three trials (AAA), a 2nd group received name, random and attribute organization for trials 1-3 respectively (NRA), a 3rd group received name organization on all three trials (NNN), and a 4th group received attribute, random, and name organization for trials 1-3 respectively (ARN). An attribute organization presents the material column-by-column from Table 1; a name organization

---

Insert Table 1 About Here

---

presents the material row-by-row. Each column or row forms a paragraph. In the random organization, the order of presentation of the cells (sentences) is random.

A different topic was presented on each trial and order of the three topics was counterbalanced over Ss. Ss were instructed that they would read three sets (i.e., three trials) of paragraphs in which the names used were nonsensical (CVC). They were told they would be given 20 seconds to read each page (paragraph) and then would be asked to recall in sentence form, after reading each set (trial), as much as they could remember.

Table 1

Topic Matrices

NEIGHBORHOOD DOGS

	eats	sleeps	chases	looks	tricks
Yom	liver	den	squirrels	furry	sit up
Dol	steak	basement	birds	big	fetches stick
Mab	ham	living room	rabbits	fat	play dead
Jav	lamb	kitchen	cats	long	roll over
Cag	veal	bedroom	mice	small	shake hands

BRANDS OF BEANS

	fortified	harvested	grows	color	transported
Pid	phosph.	Nov.	deserts	tan	car
Zum	niacin	Sept.	canyons	yellow	plane
Sek	iron	March	plateau	black	rail
Fow	potas.	June	valleys	brown	ship
Lut	calcium	Aug.	hills	red	truck

TYPES OF CANDY

	shape	eaten by	ingred.	sold in	processing time
Gif	oval	infants	sugar	super markets	hours
Weg	triangle	adults	carmel	restaur.	weeks
Tib	circular	children	cinnamon	gift shops	months
Bem	rectang.	old people	molasses	dept. stores	days
Nup	square	teenagers	choc.	drug stores	minute

During the reading, Ss were not allowed to move ahead to the next paragraph or turn back to the preceding one. After reading a set of paragraphs, Ss were required to count backwards by threes on their protocol sheets for 30 seconds. Following this they had six minutes to recall as much specific information about the preceding passage as they could.

Results

Free Recall. The results are presented in Figure 2. Analysis of variance resulted in a significant trials main effect, $F(2,96) = 5.66$, $p < .001$, and a trials by same-changing interaction, $F(2,96) = 5.33$, $p < .001$. This interaction appears to be largely due to the fact that only the AAA group shows monotonic improvement over the three trials. To further investigate this inference, we carried out a set of interaction

---

Insert Figure 2 About Here

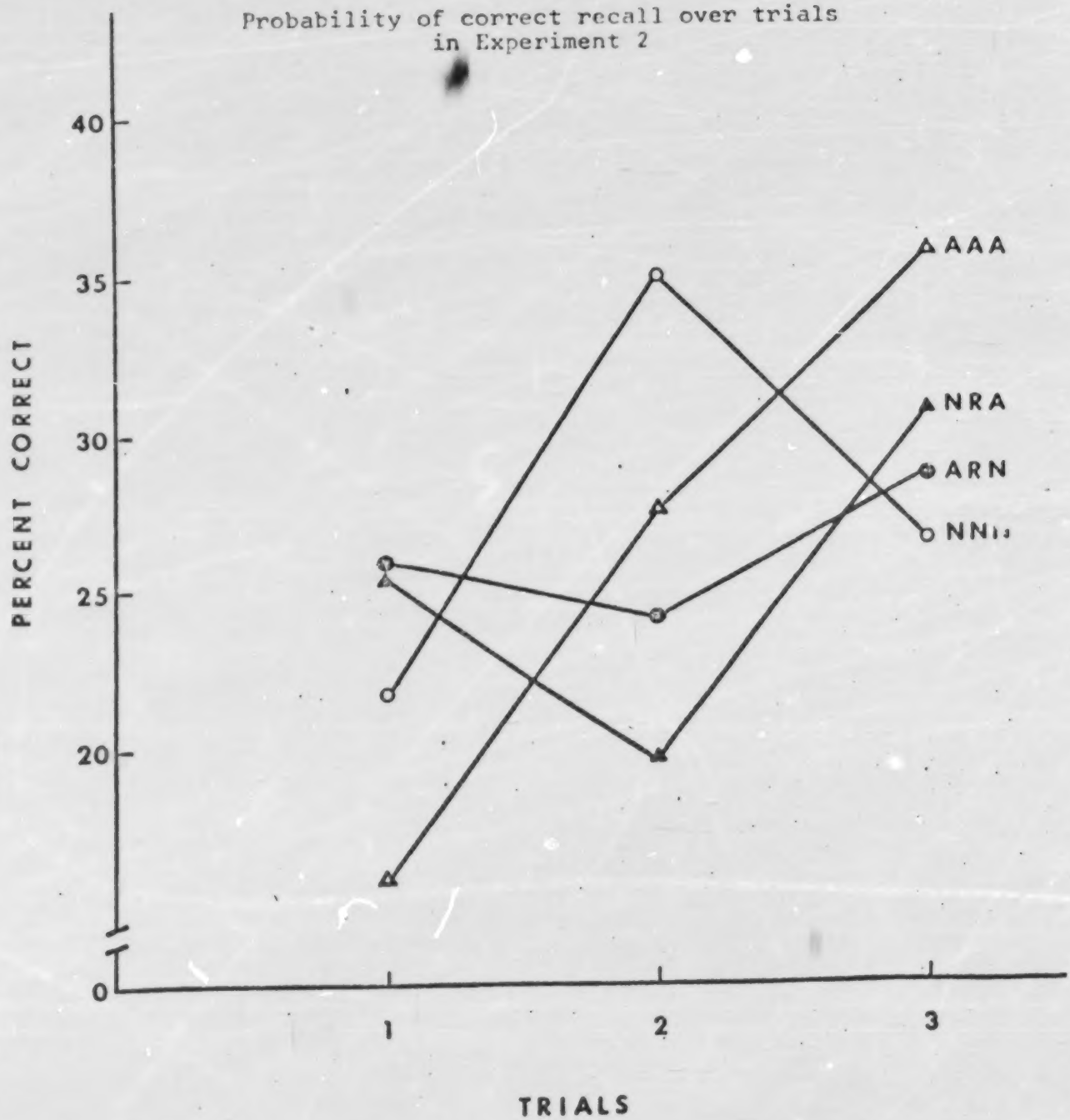
---

contrasts. The difference between NNN and AAA means did vary significantly over trials ($p < .05$) as did also the difference between the AAA mean and the mean for the other three groups combined ($p < .05$). The difference between the NNN mean and the mean for the combined NRA and ARN groups also varied significantly over trials ($p < .01$).

Intrusion Errors. Intrusion errors on trial 3 were counted. No intrusion errors were found among the attribute values; however, name intrusion errors were found. A CVC name was considered an intrusion error if a name from trials one or

Figure 2

Probability of correct recall over trials
in Experiment 2



two was substituted in trial three or if the beginning letter or last two letters came from a CVC used previously in trials one or two. When the last trial was attribute-organized there were 2 intrusion errors and when the last trial was name-organized there were 9 intrusion errors.

Discussion

In Experiment 1 the results could have been due to either a warm-up effect or a learning-to-learn effect (i.e., improved usage of a plan for retrieval). However, in Experiment 2 the presence of a significant trials by same-changing interaction clearly indicates a learning-to-learn effect. That is, different retention curves over trials as a function of organization argues against a warm-up effect. Furthermore, the average of those groups given the opportunity to use the same retrieval scheme on all three trials (AAA and NNN) improves markedly over trials, whereas this is not so for the "change" groups.

However, mere constancy of organization over trials is not a sufficient condition for improvement. Figure 2 indicates that while the AAA condition is monotonically increasing, the NNN condition shows an improvement on trial two and then a sharp decrease in retention on trial three. Myers et al. (1973) argue that in the attribute condition, Ss learn the material in semantically categorized serial lists. For example, an S would store the attribute "where grown" and then the five related attribute values: hills, valleys, plateaus, deserts, canyons. However, for the name-organized material, neither clustering or

serial cues are present; Ss learn five unrelated "lists", each consisting of a name and five semantically unrelated attribute values.

Their argument and the intrusion error evidence of Experiment 2 suggests a reason for the observed decrease in recall on trial three for the NNN group but not for the AAA group. In a transfer paradigm, as used here, the names change from trial to trial. Each trial introduces five new names. If the names are not incorporated into memory in a meaningful way in the NNN condition, as Myers et al. (1973) suggest, then these names are susceptible to interference. However, in the attribute condition the names are linked serially to the attribute categories (Myers et al., 1973) and, therefore, would not be susceptible to interference. The fact that on trial three the NNN group had five intrusion errors (total for name-organization on trial three is nine), while the AAA group had only one (total for attribute-organization on trial three is two), supports our interpretation of the observed decrease on trial three in the NNN group.

Lastly, examining the AAA and NNN retention curves in Experiment 2 (see Figure 2), the characteristic superiority of attribute over name is not exhibited until trial three. However, everywhere else in Figure 2 $\hat{A}$ is superior to N, as it is on all trials in Experiment 1 and in the research of Myers et al. (1973) and Friedman and Geitzer (1973). Moreover, the important result is that AAA monotonically increased over trials

whereas NNN did not. The point is that while both the AAA and NNN groups may learn a plan for retrieval, the attribute organization provides a better plan than name organization.

In summary, several experiments with a variety of materials indicate attribute organization is superior to name organization. In addition, Experiment 2 of this paper indicates the importance of constancy of organization over different sets of materials and that the potential for acquiring an efficient retrieval strategy under constant organization is greater under constant A than constant N organization.

References

Archer, J. E. A re-evaluation of the meaningfulness of all possible CVC trigrams. Psychological Monographs: General and Applied, 1960, 74, No. 10, Whole No. 497.

Friedman, M. P., & Greitzer, F. L. Organization and study time in learning from reading. Journal of Educational Psychology, 1972, 63, 609-616.

Myers, J. L., Pezdek, K., & Coulson, D. Effects of prose organization upon recall. Journal of Educational Psychology, in press.

Footnotes

[1]This research was supported by OE grant #OEG-1-71-0109(508).

[2]Reprint requests should be sent to Douglas Coulson, Department of Psychology, University of Massachusetts, Amherst, Mass. 01002.

The Role of Comprehension in Learning Concrete and Abstract Sentences

Kathy Pezdek and James M. Royer

The purpose of the present study was to assess the effect of comprehension on the recognition of meaning and wording changes with concrete and abstract sentences. One group was presented the sentences embedded in a context paragraph designed to increase comprehension. Recognition for meaning changes in abstract sentences was significantly higher for the sentence-embedded group than for a group presented the sentences without the paragraphs. There was no appreciable differences between the groups in recognition for wording changes in abstract sentences, nor in recognition for both meaning changes and wording changes in concrete sentences. The results of the experiment were discussed in light of recent models which propose different storage mechanisms for concrete and abstract sentences.

Begg and Pavio (1969) found that Ss presented with concrete sentences were able to detect subsequent changes in meaning in the sentences better than changes in wording. In contrast, with abstract sentences, wording changes were detected with greater facility than were changes in meaning. The interpretation given to these findings was that concrete sentences are stored as visual images whereas abstract sentences are stored as verbal strings. Johnson, Bransford, Nyberg, & Cleary (1972) have argued that the comparatively poor detection of meaning changes for abstract sentences reflects difficulties in S's comprehension of the sentences rather than providing conclusive evidence for a differential storage hypothesis. They demonstrated that the abstract sentences used in the Begg & Pavio (1969) study were more difficult to comprehend than the concrete sentences, and that the meaning change rule applied to the sentences (i.e., subject-object reversals) changed the meaning less for the abstract than for concrete sentences.

The intent of the present study is to demonstrate that the detection of meaning changes in abstract sentences can be increased by providing Ss with a treatment designed to increase comprehension of the sentences. One group of Ss in the experiment was presented the sentences embedded in a context paragraph designed to make the embedded sentences more comprehensible. Two control groups were presented the same sentences without the paragraph context. The expectation was that

the group receiving the paragraph embedded sentences would detect meaning changes in abstract sentences better than the control groups lacking the paragraph contexts.

Experiment I

METHOD

Subjects and Materials

One hundred and twenty undergraduate students from the University of Massachusetts served as subjects. They were run in groups ranging in size from five to twenty subjects.

Sixteen abstract (A) and 16 concrete (C) sentences of a constant structure were presented to the Ss. This structure was, "The (adjective) (noun) (past tense verb) a(n) (adjective) (noun)." (e.g., "The alternative version modified an established custom.") Words in the abstract and concrete sentences were equated for frequency on the basis of the Thorndike-Lorge (1944) word count. The imagery level of the nouns was evaluated when possible, based on imagery norms (Pavio, Yuille & Madigan, 1968). On the imagery scale of one to seven, the mean rating of the nouns of concrete sentences which had been formerly rated was 6.13 (11 sentences). The mean rating of the nouns of abstract sentences was 3.23 (11 sentences). Since all of the nouns used had not been rated by Pavio, Yuille & Madigan, it is important to note that these figures do not exactly represent the imagery level of the entire set of experimental sentences.

The paragraph-embedded group in the study was presented the sentences embedded in a context paragraph consisting of three sentences of similar length. The last sentence in the paragraph was always the experimental sentence. For example:

The woman on the committee refused to pass the bill until a phrase pertaining to woman's rights was included in it. The committee had never heard such a demand, but was forced to reword the bill. THE ALTERNATIVE VERSION MODIFIED AN ESTABLISHED CUSTOM.

The sentences and the context paragraphs are presented in appendix A of this paper.

Two types of changes were applied to the original sentences to produce the test sentences. Meaning changes were produced by interchanging the subject-noun and the object-noun in the sentences. The second type of change was in wording and involved substituting a synonym, matched with the original for frequency and imagery rating (when possible), for the subject-noun and leaving the rest of the sentence unchanged. These change rules were the same as those used in the Begg & Pavio study. For each sentence, meaning changes and wording changes were equally plausible. This was determined in a pilot study in which subjects rated all original sentences and test sentences. Sentences not unanamously rated as "sensible and plausible" were rejected.

The list of 32 sentences was subdivided into four sets of eight sentences. Within each set, four abstract and four concrete sentences were randomly arranged with the limitation that not more than two sentences of either type could occur in sequence. Each set of eight sentences was recorded on audio tape. A test session consisting of audio tape presentation of the changed or original sentences followed the presentation of each set of original sentences. Each

test set of eight sentences included two sentences with wording changes (W), two sentences with meaning changes (M) and four original sentences (O), arranged so that abstract and concrete sentences were equally represented in each type of test. This process of ordering sentences and determining the type of test to be applied to each sentence was carried out twice to arrive at different sequences for $Order_1$ and $Order_2$.

Design

Twenty subjects were randomly assigned to each condition in a two by three factorial design with between-subjects variables of list order (O_1 and O_2) and treatment groups (E, C_1, C_2). Additional within-subject variables of type of test sentence (W, M, or O) and level of concreteness of sentence (A or C) were included. The independent variable of particular interest was type of treatment. The experimental group (E) listened to a short paragraph providing a context to each test sentence. Each paragraph was presented within a 15 second time interval. The first control group (C_1) was allowed 15 seconds following each presented sentence to repeat and study the presented sentence. This control allowed for a comparison of conditions within this study. The second group (C_2) had an inter-sentence interval of five seconds to repeat and study the presented sentence. This control group was included to allow a comparison between the Begg & Pavio study and the present study (Begg & Pavio had on inter-sentence interval of 5 seconds).

Procedure

The experiment consisted of a familiarization task and an experimental task. In the familiarization task the Ss were first instructed

as to the requirements of the task and then presented with a practice set of six sentences. The practice sentences were generated in the same manner as the experimental sentences. The familiarization task consisted of two phases, an acquisition phase followed by a recognition phase. In the acquisition phase, four abstract and four concrete sentences were presented on audio tape. Immediately after the presentation of the last sentence in the set, the Ss were presented on audio tape a recognition test consisting of meaning or wording transforms of the original sentences or the original sentences. The task of the subjectswas to listen to each test sentence, decide if the test sentence was "identical" to the originally presented sentence or "changed", and mark the corresponding space on a response protocol sheet. In addition, subjects were asked to rate their confidence in making each response on a five-point scale. The subjects had seven seconds to respond to each test sentence.

The procedures in the experimental task were identical to those used in the familiarization session. Four trial blocks, each consisting of an acquisition phase with eight sentences, followed by a recognition test on those eight sentences, were presented on tape in succession without interruption.

RESULTS

The analyses were conducted with several purposes in mind: 1) To assess the generalizability of the Begg & Paivio results under the conditions of the present experiment; 2) To assess the comparability of the two control groups in the experiment; and 3) to examine the effects of embedding the test sentences in the context paragraphs.

Two types of analyses were used to accomplish these purposes. In the first, the dependent variable was the proportion of correctly identified changed sentences or correctly identified identical sentences. The second type of analysis utilized both sensitivity (d') scores and "cutoff" values as dependent variables (Coombs, Dawes, & Tversky, 1970). Since the analyses of the signal detection values indicated that the treatments differentially effected the bias and sensitivity scores, these analyses will be of primary interest. Reference will be made to the proportion correct analysis when it is pertinent. It should be noted that there is some ambiguity involved in analyzing the signal detection parameters due to an arbituary decision which was necessary; namely, how to treat scores of 100% correct and 0% correct. The decision was to assign 100% correct Ss a Z score of +3 and 0% correct scores a Z score of -3. Because of this decision there was some distortion of the data. For example, the d' values computed from the mean proportion correct data were generally smaller than the same values computed by summing each Ss d' score and finding the average. This distortion did not, however, alter the pattern of outcomes in the experiment. Plots of the two ways of determining d' values appeared almost identical in direction and magnitude of differences.

Comparison with Begg & Pavio (1969) The proportion of hits results from the C_2 control group and the results from the Begg & Pavio experiment are presented in Figure 1 for comparison. It is apparent from examining the figure that the pattern of outcomes from the C_2 group is different from the Begg & Pavio study in that there was an effect for type of test sentence in the present study. That is, mean-

ing changes were recognized at a higher rate than wording changes for both concrete and abstract sentences. No such effect is present in the Begg & Pavio study. It is important to note, however, that the difference of differences between recognition of meaning changes and wording changes for abstract and concrete sentences was about the same for the two studies under consideration (28% for the Begg & Pavio study and 25% for the present study).

- - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - -

Comparison of the control groups. The C_2 group in the present study differed from group E in two respects: presentation time (5 vs. 15 sec.) and absence of the paragraph context. In contrast, the C_1 group differed from group E only in absence of the paragraph context. Because of these differences a preliminary analysis was performed on the two control groups to assess their comparability. A 2 (C_1 vs. C_2) x 2 (A vs. C sentences) x 2 (meaning vs. wording changes) x 2 (O_1 vs. O_2) analysis of variance on d' scores indicated a main effect for sentence type, $F(1,76) = 21.2$, $p < .01$, with concrete sentences having a higher mean d' (2.74) than abstract sentences (mean d' = 1.84). Type of sentence change was also a reliable effect, $F(1,76) = 73.7$, $p < .01$ with meaning changes being recognized with greater sensitivity (Mean d' = 2.89) than wording changes (mean d' = 1.69). In addition to the main effects, there was an interaction between sentence type and test change type, $F(1,76) = 27.5$, $p < .01$, such that the difference between meaning and wording changes was greater for concrete sentences than for abstract sentences. Two second order interactions were also noted in the analysis. Both the groups x order x test type,

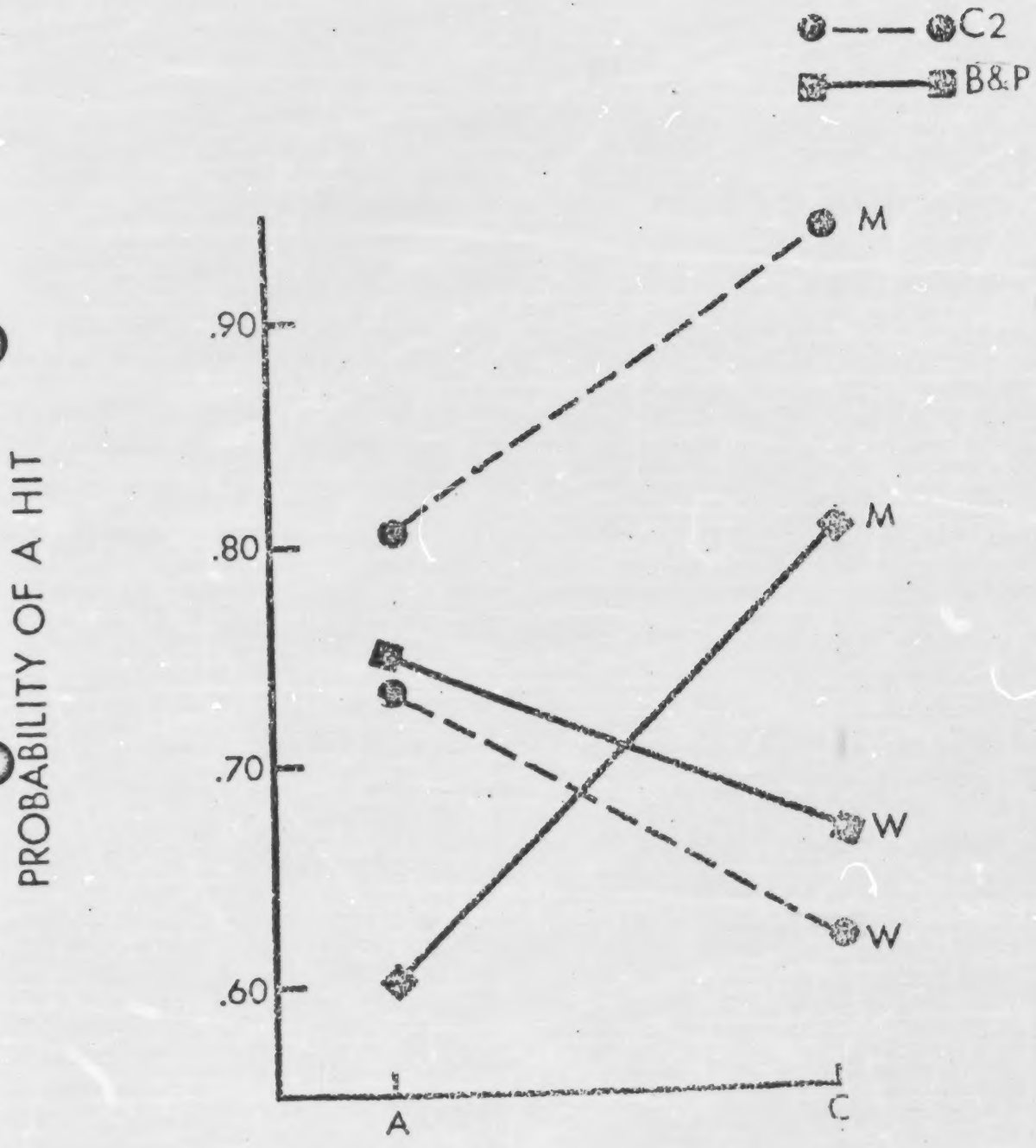


Figure 1. Comparison of P(hit) on meaning and wording changes as a function of sentence concreteness for Begg and Paivio's data and Control.

$F(1,76) = 8.7$, $p < .01$, interactions were significant sources of variance.

The same analysis performed on response cutoff scores indicated that the main effect for treatment groups was not significant. However, a significant main effect of list order, $F(1,76) = 8.4$, $p < .01$, was obtained with response cutoff scores. No other results of interest to the study were obtained in analysis using cutoff scores as the dependent measure. A complete listing of all of the analyses performed are presented in Appendix B of this paper.

Effect of Paragraph Context. Since the analyses comparing groups C_1 and C_2 suggested that the two groups yielded essentially the same experimental outcomes (i.e., the group factor entered into only one second order interaction), the effect of embedding the treatment sentences in the paragraph contexts was assessed through a comparison of groups E and C_1.

The analysis included the same factors assessed in the comparison of groups C_1 and C_2. The analysis of the d' scores is of primary interest in the study. The pattern of outcomes is presented in Figure 2. There was two significant main effects in the analysis. Abstract sentences ($\bar{X} = 1.95$) produced lower d' socres than did concrete sentences ($\bar{X} = 2.53$), $F(1,76) = 11.8$, $p < .01$. Type of test also produced a significant effect with Ss responding with greater sensitivity to meaning changes ($\underline{X} = 2.92$) than to wording changes ($\bar{X} = 1.55$), $F(1,76) = 83.3$, $p < .01$. In addition to the main effects there were two significant interactions. The groups X sentence type interaction, $F(1,76) = 5.46$, $p < .05$, was reliable such that there was a greater difference between groups E and C_1 ($\bar{X} = 2.29$ and 2.76

respectively) with concrete sentences than there was with abstract sentences ($\bar{X}$ = 2.11 and 1.79, respectively). The sentence type X test type interaction was also significant, $F(1,76) = 7.27$, $p < .01$, with the difference between meaning changes and wording changes being greater for concrete sentences ($\bar{X}$ = 1.66 and 3.40), respectively) than for abstract sentences ($\bar{X}$ = 1.45 and 2.45, respectively). The critical second order interaction of groups X sentence type X test type was marginally significant, $F(1,76) = 3.47$, $p = .06$. A simple effects test for the predicted increase in sensitivity for meaning changes with abstract sentences due to embedding the sentences in the paragraph context was significant, one tailed $t(76) = 1.93$, $p < .05$. The means for the critical comparison of d' scores for meaning changes in abstract sentences for groups E and C_1 were, respectively 2.82 and 2.08.

- - - - - - - - - - - - -

Insert Figure 2 about here

- - - - - - - - - - - - -

It should be noted that the analysis using proportion of hits as a dependent variable produced the same pattern of outcomes and did yield a significant second order interaction (groups X meaning changes X sentence type), $F(1,76) = 4.66$, $p < .05$. The remaining outcomes in this analysis were essentially the same as those for the d' analysis, with one exception. Order effects contributed to one significant first order interaction (order X sentence type) and one second order interaction (order X sentence type X test type). The fact that order did not contribute to any reliable effects in the d' analysis, but did contribute to the significant order X sentence type interaction in

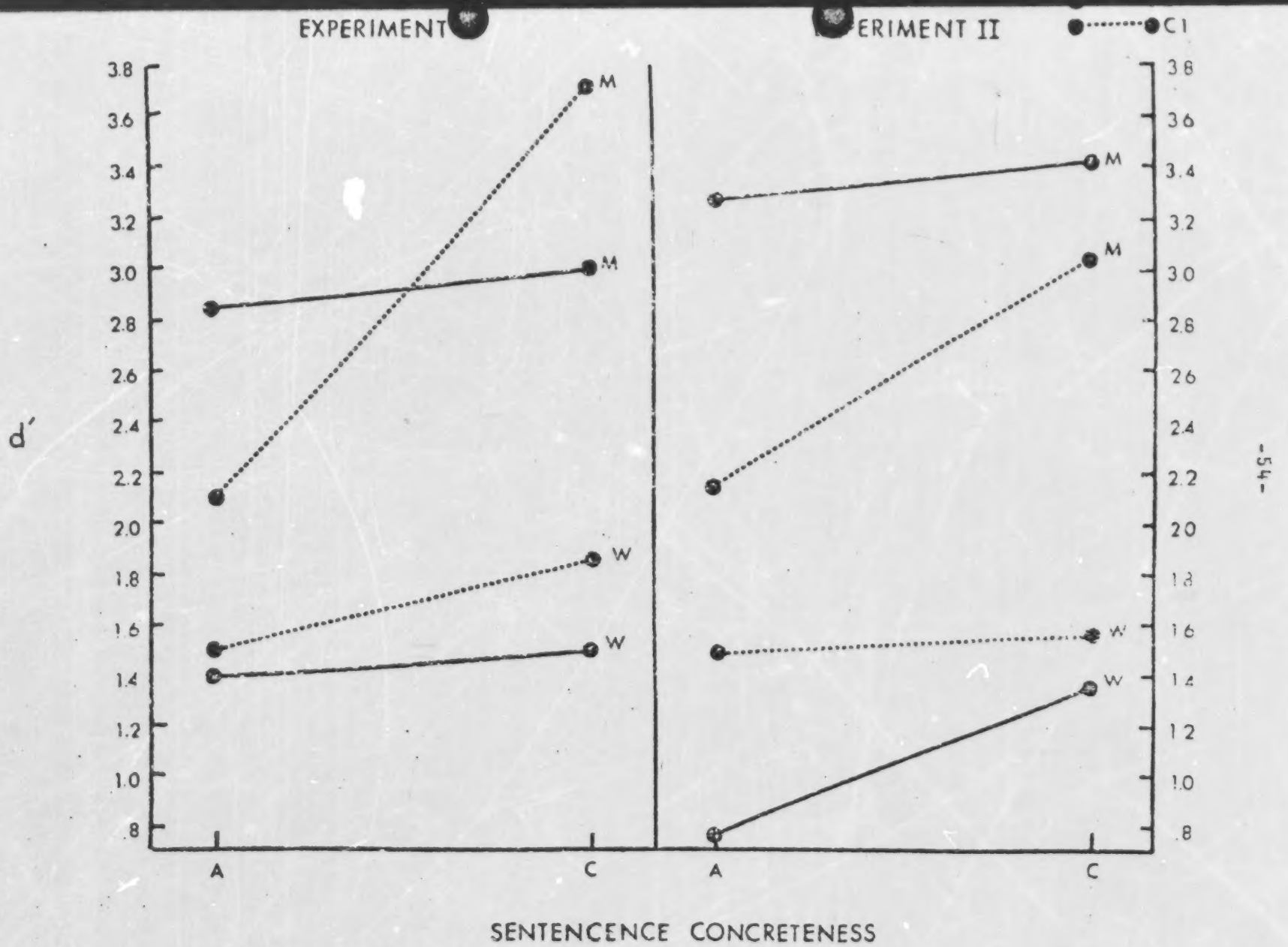


Figure 2. Comparison of d' values obtained in Experiments I and II of the present study for groups E and Control$_2$.

the cutoff value analysis suggests that the two orders of sentences produced different response biases but did not influence the sensitivity with which the Ss made their decision.

Hit Rate for Original Test Sentences and Confidence Ratings. A separate 2 (A vs. C sentences) X 2 (meaning vs. wording changes) X 3 (groups E, C_1, C_2) X 2 ($order_1$ vs. $order_2$) analysis of the hit rate for original sentences yielded a significant main effect for sentence concreteness, $F(1,114) = 27.3$, $p < .01$, with original concrete sentences ($\bar{X} = .779$) being recognized with greater accuracy than original abstract sentences ($\bar{X} = .681$). In addition, there was a significant effect for order $F(1,114) = 5.93$, $p = .01$, with O_1 original sentences being recognized with greater accuracy than O_2 original sentences. A significant groups X concreteness interaction, $F(2,114) = 9.76$, $p < .01$, was also noted with the difference between group E and the two control groups (which were almost identical) being smaller with abstract sentences than with conceret sentences.

There were several significant outcomes in the analysis of the confidence rating data but only the most pertinent ones will be reported. There were confidence rating differrences between the three types of test items (meaning, wording, original). $F(2,228) = 47$, $p < .01$. On a five point scale original items were rated with the least confidence ($\bar{X} = 3.69$) followed by wording changed items ($\bar{X} = 3.91$), and with meaning changed items being recognized with the greatest confidence ($\bar{X} = 4.17$). In addition, concrete items were rated with greater confidence than abstract items, $f(1,114) = 199.5$, $p < .01$.

A somewhat surprising outcome was the lack of an effect for treatment groups, $F < 1$. The confidence of Ss did not vary systematically due to either extending the sentence presentation duration or to embedding

the sentences in the context-paragraph.

Experiment II

Since the predicted increase in detection of meaning changes in abstract sentences as a function of paragraph embedding was only marginally supported in Experiment I, we attempted to obtain a replication of the effect in a second experiment. The control group comparable to the Begg & Pavio group (C_2) was eliminated from the replication experiment because of the comparability established between the C_1 and C_2 groups in Experiment 1.

Method

Twenty undergraduate students were randomly assigned to each condition in a 2 (groups E and C_1) X 2 (List order O_1 and O_2) factorial design. The material, within-subject variables, and procedures were identical to those used in Experiment I. The only difference between the two experiments was a change of experimenters and the absence of the C_2 group from Experiment II.

Results

The analysis of d' scores was of primary interest in the experiment. In particular, attention was focused on the predicted interaction between groups, sentence type, and test type. This interaction is displayed in the second panel of Figure 2. Whereas the interaction was only marginally significant in Experiment I, the effect was much more substantial in Experiment II, $F(1,76) = 7.9$, $p < .01$. The predicted increase in detection sensitivity for meaning changes with abstract sentences as a function of paragraph embedding was assessed using a simple-effects test. Again the effect was more substantial ($t(76) = 3.7$, $p < .01$) than in the first experiment.

The remainder of the outcomes were essentially the same as those for Experiment I. This was established by combining the data from both experiments in an overall analysis with experiments as a factor in the analysis. The experiments factor did not contribute to any significant outcomes in the analysis. The ANOVA tables using d' values as a dependent variable for Experiment II and for Experiments I and II combined are reported in Appendix B.

DISCUSSION

The major purpose of the study was to examine the effect on the sentence recognition due to embedding abstract and concrete sentences in a paragraph context. Begg and Pavio (1969) have taken the strong position that concrete sentences are stored in memory primarily as images whereas abstract sentences are stored primarily as words. Several lines of evidence suggest that the Begg & Pavio results stem from inadequate comprehension of the abstract sentences (Johnson, et al., 1972).

Presumably, abstract sentences of the type used in the present study are more difficult to comprehend than concrete sentences. This likelihood might lead Ss to a strategy whereby they simply attempt to memorize the word strings when presented with abstract sentences. Our paragraph-embedding treatment was an attempt to increase the comprehension of abstract sentences and thereby increase the likelihood that changes in the meaning of the sentences would be detected.

The visual comparison presented in Figure 1 of the outcomes of our C_2 group and the Begg and Pavio (1969) study suggest one major difference in the results: i.e., the greater extent to which Ss detected meaning changes in the C_2 group at both levels of sentence

abstractness. This difference is probably due to differences in the sentences used in the two studies. Our materials were based on the sentences used in the Begg & Pavio study, but the design differences between the two studies required that changes be made and that additional sentences be generated.[3] It is probably the case that our final pool of materials consisted of sentences in which meaning changes were more easily detected than in the sentences used by Begg & Pavio. It is important to emphasize however, the very prominent interaction between sentence abstractness and test type that exists in both studies.

The comparison of the C_1 and C_2 groups in the study indicated that there were no appreciable differences between the two groups due to extending the presentation time for each sentence an additional 10 seconds (i.e., 5 vs. 15 sec.). This suggests that if comprehension is an important determinant of the ability to detect meaning changes in sentences, simply extending the time each sentence is presented has little effect on the degree to which Ss comprehend the sentences.

The null effects of extending presentation time can be contrasted with the effects due to embedding the sentences in the paragraph context (see Figure 2). The paragraph-embedding treatment had virtually no effect on responses to wording changes. However, the treatment did effect the accuracy with which Ss recognized meaning changes. Specifically, the combined results of Experiments I and II offer strong support for the conclusion that embedding the abstract sentences in paragraphs substantially improves a subject's ability to detect meaning changes in the sentences.

References

Begg, I., & Pavio, A. Concreteness and imagery in sentence meaning. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 821-827.

Coombs, C. H., Dawes, R. M., & Tversky, A. Mathematical Psychology, An Elementary Introduction. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970.

Johnson, M. K., Bransford, J. D., Nyberg, S. E., & Cleary, J. J. Comprehension factors in interpreting memory for abstract and concrete sentences. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 451-454.

Pavio, A., Yuille, J. C., & Madigan, S. A. Concretness, imagery, and meaningfulness values for 925 nouns. Journal of Experimental Psychology, 1968, 76, Monograph Supplement No. 1, Part 2.

Thorndike, E. L., & Lorge, I. The teacher's workbook of 30,000 words. New York: Teachers College Press, 1944.

Footnotes

[1]This paper is based upon an M.S. thesis completed by the first author under the direction of the second.

[2]During the completion of the study the first author was supported by Office of Education grant OEG-1-71-0109(508). The final version of the paper was completed while the second author was supported by NSF Development grant GU4041. We would like to thank Gary Evans, Jerome Myers, Charles Clifton, and Jane Perlmutter for a critical reading of an early version of this paper.

[3]In the present study it was necessary to utilize more test sentences than the number used by Begg and Pavio, 1969. The primary reason for this was design differences between the two studies. In the present study, concreteness and test type were within subjects' variables, whereas these were between subjects variables in the Begg and Pavio study. Also, Begg and Pavio used "filler sets" (sentences which could only be transformed to one of the two test type sentences) which were not included in the analyses. All sentences in the present study were relevant to the analyses.

Appendix A

Presentation Sentences Used in the Present Study with Wording (W) and Meanings (M) Test Sentences

Abstract Sentences

1. The final reason supplied the adequate motivation.
W. The final consideration supplied the adequate motivation.
M. The final motivation supplied the adequate reason.

2. The solemn congregation encouraged an impressive service.
W. The solemn gathering encouraged an impressive service.
M. The solemn service encouraged an impressive congregation.

3. The foreign faith aroused an enduring interest.
W. The foreign belief aroused an enduring interest.
M. The foreign interest aroused an enduring faith.

4. The introductory statement promised a logical treatment.
W. The introductory assertion promised a logical greatment.
M. The introductory treatment promised a logical statement.

5. The simple mechanism reflected an outdated technology.
W. The simple machinery reflected an outdated technology.
M. The simple technology reflected an outdated mechanism.

6. The vague concern survived a renewed enthusiasm.
W. The vague opinion survived a renewed enthusiasm.
M. The vague enthusiasm survived a renewed concern

7. The complex review revealed an objective position.
W. The complex summary revealed an objective position.
M. The complex position revealed an objective review.

8. The strange situation altered the accepted conclusion.
W. The strange circumstance altered the accepted conclusion.
M. The strange conclusion altered the accepted situation.

9. The unpleasant factor caused a dismal silence.
W. The unpleasant element caused a dismal silence.
M. The unpleasant silence caused a dismal factor.

10. The complicated proof explained a limited formula.
W. The complicated calculation explained a limited formula.
M. The complicated formula explained a limited proof.

11. The original condition implied an essential balance.
W. The original provision implied an essential balance.
M. The original balance implied an essential condition.

12. The recent pact made an approved solution.
W. The recent accord made an approved solution.
M. The recent solution made an approved pact.

13. The required duty involved a standard payment.
W. The required task involved a standard payment.
M. The required payment involved a standard duty.

14. The ridiculous mistake nullifed a prior commitment.
W. The ridiculous error nullified a prior committment.
M. The ridiculous commitment nullified a prior mistake.

15. The alternative version modified an established custom.
W. The alternative copy modified an established custom.
M. The alternative custom modified an established version.

16. The actual quotation lacked a rational idea.
W. The actual expression lacked a rational idea.
M. The actual idea lacked a rational quotation.

Concrete Sentences

1. The hungry boar attacked a sleeping coyote.
W. The hungry hog attacked a sleeping coyote.
M. The hungry coyote attacked a sleeping boar.

2. The pompous monarch confronted the triumphant queen.
W. The pompous king confronted the triumphant queen.
M. The pompous queen confronted the triumphant monarch.

3. The hollow tree housed an old tomb.
W. The hollow trunk housed an old tomb.
M. The hollow tomb housed an old tree.

4. The unfamiliar helper accused a reckless prisoner.
W. The unfamiliar assistant accused a reckless prisoner.
M. The unfamiliar prisoner accused a reckless helper.

5. The carefree merchant annoyed the timid woman.
W. The carefree dealer annoyed the timid woman.
M. The carefree woman annoyed the timid merchant.

6. The white truck passed a rickety automobile.
W. The white van passed a rickety automobile.
M. The white automobile passed a rickety truck.

7. The young author cherished the homely girl.
W. The young writer cherished the homely girls.
M. The young girl cherished the homely author.

8. The crippled forger killed the tortured slave.
W. The crippled blacksmith killed the tortured slave.
M. The crippled slave killed the tortured forger.

9. The enthusiastic professor welcomed the familiar doorman.
W. The enthusiastic instructor welcomed the familiar doorman.
M. The enthusiastic doorman welcomed the familiar professor.

10. The delicate maiden watched the great dreamer.
W. The delicate damsel watched the great dreamer.
M. The delicate dreamer watched the great maiden.

11. The strong painter bullied the seedy beggar.
W. The strong artist bullied the seedy beggar.
M. The strong beggar bullied the seedy painter.

12. The overgrown stalk shaded a delicate mushroom.
W. The overgrown stem shaded a delicate mushroom.
M. The overgrown mushroom shaded a delicate stalk.

13. The polite servant introduced an intolerant doctor.
W. The polite slave introduced an intolerant doctor.
M. The polite doctor introduced an intolerant servant.

14. The smooth rock struck a heavy pot.
W. The smooth stone struck a heavy pot.
M. The smooth pot struck a heavy rock.

15. The poor scoundrel called a sluggish policeman.
W. The poor villain called a sluggish policeman.
M. The poor policeman called a sluggish scoundrel.

16. The alert laborer pursued the talkative student.
W. The alert worker pursued the talkative student.
M. The alert student pursued the talkative laborer.

Contextual Material Presented to the Experimental Treatment Group

Wild animals seldom abuse territorial boundaries except for self-defense. The need for food can explain many aggressive acts which would not otherwise occur. THE HUNGRY BOAR ATTACKED A SLEEPING COYOTE.

After much arguing, the young child could still not understand why he could not play in the mud. When his mother threatened punishment, he changed his mind. THE FINAL REASON SUPPLIED THE ADEQUATE MOTIVATION.

The king had never felt threatened by another man, not to mention a woman. The victorious queen tried to avoid him on her trip through his kingdom, but he found a way to talk to her anyway. THE POMPOUS MONARCH CONFRONTED THE TRIUMPHANT QUEEN.

The woodsman had died long ago in the middle of a severe winter. Because the ground was too hard to bury him, his body was placed inside a huge rotted tree. THE HOLLOW TREE HOUSED AN OLD TOMB.

The minister had always had a difficult time attracting the town's people to church. The night of the assasinated governor's funeral was an exception. THE SOLEMN CONGREGATION ENCOURAGED AN IMPRESSIVE SERVICE.

The foreign-exchange student from India spoke at an attentive high school assembly Wednesday. Much of her talk revealed the fact that Buddhism was a major guiding force in her life. THE FOREIGN FAITH AROUSED AN ENDURING INTEREST.

Because the convict frequently broke tools which he borrowed from the prison shop, he often was blamed for others' mistakes. A new prison aid found that using the man as a scape-goat for his own short-sightedness was often profitable. THE UNFAMILIAR HELPER ACCUSED A RECKLESS PRISONER.

The president was anxious to study the committee's review of the welfare program. They had outlined their proposal to him that morning. THE INTRODUCTORY STATEMENT PROMISED A LOGICAL TREATMENT.

The first job of the Peach Corp engineer was to replace the horse-driven well pump with a small electric pump. The old type pump was too slow and was holding up irrigation of the field. THE SIMPLE MECHANISM REFLECTED AN OUTDATED TECHNOLOGY.

The jolly man was selling fruit at the market at the booth usually occupied by the butcher. He only laughed when one quiet but obviously disoriented shopper scurried around in front of him. THE CAREFREE MERCHANT ANNOYED THE TIMID WOMAN.

For months no one really took a stand on how they felt about environmental legislation. When a sanitary land-fill area was planned for the outskirts of the city, everyone was up in arms. THE VAGUE CONCERN SURVIVED A RENEWED ENTHUSIASM.

Driving along the turnpike is particularly trying when you are stuck behind a slow car. A truck was anxious for a chance to pull into the passing lane. THE WHITE TRUCK PASSED A RICKETY AUTOMOBILE

The jury tried to express their feelings without showing signs of emotional involvement in the case. After a one hour oration, they stated their decision. THE COMPLEX REVIEW REVEALED AN OBJECTIVE POSITION.

The gentleman was noted for the compassion he revealed in his novels. That is why no one doubted that he would marry the poor maiden whom he loved. THE YOUNG AUTHOR CHERISED THE HOMELY GIRL.

After weeks of debate, the equal-rights committee voted unanamously in favor of a female president. Unexpectedly, the only eligible female was hospitalized as a result of a serious accident. THE STRANGE SITUATION ALTERED THE ACCEPTED CONCLUSION.

The brawny man could not run and had only his strength to defend himself with. The mad servant had not seen the anvil in his hand. THE CRIPPLED FORGER KILLED THE TORTURED SLAVE.

The excited ski team was suddenly quited by the news that one of their team-mates had fallen during the slalom race. This disaster would ruin any chance of a gold medal for the team. THE UNPLEASANT FACTOR CAUSED A DISMAL SILENCE.

The couple was anxious to revisit the people they had met the previous summer in London. When they pulled up to their favorite restaurant, the friendly face they had remembered was there to greet them. THE ENTHUSIASTIC PROFESSOR WELCOMED THE FAMILIAR DOORMAN.

The statistics class was convinced that they were wasting their time going through such a long proof. The formula they were deriving would not even be very useful to any of them. THE COMPLICATED PROOF EXPLAINED A LIMITED FORMULA.

Any changes that were attempted in the factory threw the whole system off. The old routine, with all of its faults, still ran smoother than any innovation plan. THE ORIGINAL CONDITION IMPLIED AN ESSENTIAL BALANCE.

The young girl was infatuated by her friend who gazed idly toward the sky. She sat silently by him and hesitated to disturb him. THE DELICATE MAIDEN WATCHED THE GREAT DREAMER.

The United States agreed to offer financial assistance to Turkey if in turn Turkey would discourage exportation of Morphine to the U. S. The U. S. did not want Morphine smuggled into this country and Turkey needed financial aid. THE RECENT PACT MADE AN APPROVED SOLUTION.

The noted master was insensitive in his treatment of others. When a helpless derelict stumbled into him and asked for money, the villager was annoyed. THE STRONG PAINTER BULLIED THE SEEDY BEGGAR.

Forest mushrooms flourish in areas which receive a moderate amount of sun light. Some underbrush which grows faster than the mushrooms stifles the growth of muchrooms. THE OVERGROWN STALK SHADED A DELICATE MUSHROOM.

The butler did not look forward to presenting the doctor to the host. Although guests were often abrupt with him, the butler performed his job admirably. THE POLITE SERVANT INTRODUCED AN INTOLERANT DOCTOR.

The importation laws required that jewelry being shipped over the border be accounted for according to its weight. All jewelry, regardless of its worth was considered on the basis of the same criterion. THE REQUIRED DUTY INVOLVED A STANDARD PAYMENT.

The teacher had promised the class a field trip as a reward for their attendance. He remembered too late that he had neglected to reserve a bus, and therefore had to cancle the trip. THE RIDICULOUS MISTAKE NULLIFIED A PRIOR COMMITMENT.

The child was not aware of the value of the utensils he had pulled out of the kitchen cabinet. He threw a homemade paperweight toward the cabinet. THE SMOOTH ROCK STRUCK A HEAVY POT.

The old man knew that he was considered to be a rather shady character and never thought that he would have to rely on civil authority for protection. One night the fellow lost his way in the city and helplessly sought assistance. THE POOR SCOUNDREL CALLED A SLUGGISH POLICEMAN.

The woman on the committee refused to pass the bill until a phrase pertaining to woman's rights was included in it. The committee had never heard such a demand, but was forced to reword the bill. THE ALTERNATIVE VERSION MODIFIED AN ESTABLISHED CUSTOM.

The union member was curious as to the views of college students on the Vietnam war. He was thrilled when he overheard a co-ed avidly discussing the topic with a friend. THE ALERT LABORER PURSUED THE TALKATIVE STUDENT.

It was popular for people to quote Mark Twain in regards to his statement on old age. As would be expected of the humorous writer, his comments on old age reflected a make-believe senility on his own part. THE ACTUAL QUOTATION LACKED A RATIONAL IDEA.

Appendix B

Analyses of variance tables and table of means for study.

Table 1

Analysis of Variance and Table of Means for P(Hit) Data for Groups E, C_1 & C_2

Source of Variance	df	MS	F
G Groups)	2	.0396	.68
O (Order)	1	.4380	8.10**
C (Concreteness)	1	.1333	2.70
T (Test Type)	1	5.1047	118.71**
GO	2	.1005	1.86
GC	2	.0145	.30
OC	1	.7521	15.26**
GT	2	.0297	.69
OT	1	.0880	2.05
CT	1	.8333	20.22**
S (GO)	114	.0541	
GOC	2	.0005	.01
GOT	2	.1911	4.44*
GCT	2	.1286	3.74*
OCT	1	.5333	15.50**
SC (GO)	114	.0493	
ST (GO)	114	.0431	
GOCT	2	.0536	1.558
SCT (GO)	114	.0344	

*Significant at .05α level
**Significant at .01α level

Table of Means

		E		C_1		C_2	
		O_1	O_2	O_1	O_2	O_1	O_2
A	Wording Charge	.700	.625	.763	.625	.713	.750
	Meaning Charge	.913	.850	.725	.825	.788	.813
C	Wording Charge	.613	.750	.550	.725	.425	.813
	Meaning Charge	.913	.963	.888	.975	.938	.938

Table 2

Analysis of Variance and Table of Means for

d' Data for Groups E, C_1 and C_2

Source of Variance	df	MS	F
G (Groups)	2	.4575	<1.00
O (Order)	1	.0459	<1.00
C (Concreteness)	1	52.1704	19.53**
T (Test Type)	1	207.0050	131.43**
GO	2	2.4941	<1.00
GC	2	6.9790	2.61
OC	1	10.8535	4.06*
GT	2	1.3879	<1.00
OT	1	1.2078	<1.00
CT	1	30.0040	21.71**
S (GO)	114	2.6620	
GOC	2	1.2928	<1.00
GOT	2	6.4728	4.11*
GCT	2	4.7410	3.43*
OCT	1	13.1420	9.51**
SC (GO)	114	2.6713	
ST (GO)	114	1.5750	
GOCT	2	.7424	<1.0
SCT (GO)	114	1.3823	

*Significant at .05 α level
**Significant at .01 α level

Table of Means

		E		C_1		C_2	
		O_1	O_2	O_1	O_2	O_1	O_2
A	Wording Charge	1.605	1.197	1.721	1.262	2.016	1.347
	Meaning Charge	2.921	2.713	1.703	2.462	2.465	1.763
C	Wording Charge	1.209	1.740	1.632	2.069	1.084	2.371
	Meaning Charge	3.175	3.056	3.450	3.900	4.034	3.370

Table 3

Analysis of Variance and Table of Means for

P(Hit) Data for Groups E and C_1

Source of Variance	df	MS	F
G (Groups)	1	.0781	1.48
O (Order)	1	.0945	1.80
C (Concreteness)	1	.1531	3.50
T (Test Type)	1	3.6125	76.05**
GO	1	.0383	<1.00
GC	1	.0031	<1.00
GC	1	.4883	11.17**
GT	1	.0500	1.05
OT	1	.0070	<1.00
CT	1	.3125	8.18**
S (GO)	76	.0526	
GOC	1	.0008	<1.00
GOT	1	.0633	1.33
GCT	1	.1531	4.01**
OCT	1	.2258	5.91*
SC (GO)	76	.0437	
ST (GO)	76	.0475	
GOCT	1	.0633	1.56
SCT (GO)	76	.0382	

*Significant at .05α-level
**Significant at .01α level

Table of Means

		E		C_1	
		O_1	O_2	O_1	O_2
A	Wording Charge	.700	.625	.763	.625
	Meaning Charge	.913	.850	.725	.825
C	Wording Charge	.613	.750	.550	.725
	Meaning Charge	.913	.963	.888	.975

Table 4

Analysis of Variance and Table of Means for

d' Data for Groups E and C_1

Source of Variance	df	MS	F
G (Groups)	1	.4249	<1.00
O (Order)	1	1.2090	<1.00
C (Concreteness)	1	26.9648	11.80**
T (Test Type)	1	149.7481	83.32**
GO	1	2.4240	<1.00
GC	1	12.4682	5.46*
OC	1	3.2629	1.43
GT	1	2.0565	1.19
OT	1	0.7635	<1.00
CT	1	10.6394	7.27**
S (GO)	76	2.6356	
GOC	1	0.2436	<1.00
GOT	1	3.5349	2.04
GCT	1	5.0851	3.47
OCT	1	5.2667	3.60
SC (GO)	76	2.2845	
ST (GO)	76	1.7349	
GOCT	1	0.1581	<1.00
SCT (GO)	76	1.4638	

*Significant at .05α level
**Significant at .01α level

Table of Means

		E		C_1	
		O_1	O_2	O_1	O_2
A	Wording Charge	1.605	1.197	1.721	1.262
	Meaning Charge	2.921	2.713	1.703	2.462
C	Wording Charge	1.209	1.740	1.632	2.069
	Meaning Charge	3.175	3.056	3.450	3.900

Table 5

Analysis of Variance and Table of Means for

P(Hit) Data for Groups C_1 and C_2

Source of Variance	df	MS	F
G (Groups)	1	.0125	<1.00
O (Order)	1	.5695	10.83**
C (Concreteness)	1	.0781	1.3971
T (Test Type)	1	2.9070	68.40**
GO	1	.0633	1.20
GC	1	.0281	<1.00
OC	1	.4883	8.74**
GT	1	.0008	<1.00
OT	1	.0781	1.84
CT	1	1.0695	33.84**
S (GO)	76	.0526	
GOC	1	.0008	<1.00
GOT	1	.3781	8.98**
GCT	1	.0070	<1.00
OCT	1	.6125	19.38**
SC (GO)	76	.0559	
ST (GO)	76	.0425	
GOCT	1	.0031	<1.00
SCT (GO)	76	.0316	

*Significant at .05 α level
**Significant at .01 α level

Table of Means

		C_1		C_2	
		O_1	O_2	O_1	O_2
A	Wording Charge	.763	.625	.713	.750
	Meaning Charge	.725	.825	.788	.813
C	Wording Charge	.550	.975	.938	.938
	Meaning Charge	.888	.975	.938	.938

Table 6

Analysis of Variance and Table of Means for

d' Data for Groups C_1 and C_2

Source of Variance	df	MS	F
G (Groups)	1	.0786	<1.00
O (Order)	1	.2413	<1.00
C (Concreteness)	1	64.2459	21.16**
T (Test Type)	1	116.3296	73.72**
GO	1	4.6883	1.72
GC	1	.5019	<1.00
OC	1	8.3228	2.74
GT	1	.0003	<1.00
OT	1	.7104	<1.00
CT	1	38.5052	27.49**
S (GO)	76	2.7299	
GOC	1	2.4716	<1.00
GOT	1	12.9368	8.20**
GCT	1	..4739	<1.00
OCT	1	12.1887	8.70**
SC (GO)	76	3.0356	
ST (GO)	76	1.5780	
GOCT	1	.6380	<1.00
SCT (GO)	76	1.4008	

*Significant at .05 α level
**Significant at .01 α level

Table of Means

		C_1		C_2	
		O_1	O_2	O_1	O_2
A	Wording Charge	1.721	1.262	2.016	1.347
	Meaning Charge	1.703	2.462	2.465	1.763
C	Wording Charge	1.632	2.069	1.084	2.371
	Meaning Charge	3.450	3.900	4.034	3.370

Table 7

Analysis of Variance and Table of Means for

d' Data for Groups E and C_1

Experiment II

Source of Variance	df	MS	F
G (Groups)	1	1.5331	.48
O (Order)	1	2.2313	.67
C (Concreteness)	1	15.0568	4.61*
T (Test Type)	1	227.1970	69.81**
GO	1	15.3374	4.59*
GC	1	0.1757	.05
OC	1	19.8358	6.08*
GT	1	28.5814	21.91**
OT	1	1.8536	1.42
CT	1	0.6339	.63
S (GO)	76	3.3380	
GOC	1	14.8803	4.56*
GOT	1	0.6987	.53
GCT	1	7.9157	7.85**
OCT	1	4.5886	4.55*
SC (GO)	76	3.2644	
ST (GO)	76	1.3043	
GOCT	1	1.2902	1.28
SCT (GO)	76	1.0089	

*Significant at .05 α level
**Significant at .01 α level

Table of Means

		E		C_1	
		O_1	O_2	O_1	O_2
A	Wording Charge	1.27	.30	1.77	1.25
	Meaning Charge	3.48	3.11	2.59	1.80
C	Wording Charge	1.45	1.34	.81	2.37
	Meaning Charge	3.94	2.97	2.66	3.50

Table 8

Analysis of Variance

d' Data for Groups E and C_1

Across Experiments I and II

Source of Variance	df	MS	F
D (Exp. I or II)	1	0.8933	.30
G (Groups)	1	0.1719	.06
O (Order)	1	0.0777	.00
C (Concreteness)	1	41.1603	14.84**
T (Test Type)	1	372.9240	245.41**
DG	1	1.7861	.60
DO	1	3.3626	1.13
GO	1	14.9781	5.02*
DC	1	0.8613	.31
GC	1	7.8022	2.81
OC	1	19.5944	7.06**
DT	1	4.0211	2.65
GT	1	22.9856	15.13**
OT	1	0.1189	.08
CT	1	8.2337	6.66**
DGO	1	2.7834	.93
DGC	1	4.8417	1.75
DOC	1	3.5043	1.26
GOC	1	5.6580	2.04
DGT	1	7.6523	5.04*
DOT	1	2.4983	1.64
GOT	1	0.5452	.36
DCT	1	3.0396	2.46
GCT	1	12.8448	10.39**
OCT	1	9.8436	7.96**
S (DGO)	152	2.9868	
DGOC	1	9.4658	3.41
DGOT	1	3.6884	2.43
DGCT	1	0.1559	.13
DOCT	1	0.0117	.01
GOCT	1	0.2726	.22
SC (DGO)	152	2.7745	
ST (DGO	152	1.5196	
DGOCT	1	1.1756	.95
SCT (DGO)	152	1.2364	

Retrieval Processes in Recall[1]

Jane Perlmutter, Patricia Sorce, and Jerome L. Myers

A reaction time (RT) paradigm was developed to study retrieval processes in paired associate (PA)recall. Prior to the experimental session Ss learned lists of PAs (varying in length from three to twenty-four pairs); during the experimental session, Ss' RT to say the response word from the onset of a visually presented stimulus word was measured. The implications of several classes of retrieval models were discussed in the context of this paradigm. The shape of the RT-list length function, practice, and sequential effects were all of interest in distinguishing among models. Four experiments were reported which were designed to 1) establish the baseline effects in this paradigm, 2) determine which of these effects should be attributed to the retrieval stage of processing, and 3) investigate the effect of semantic memory in this task. Results suggest that serial scanning models are inadequate to handle the data from this task. Strength models, on the other hand, seem to capture the qualitative effects present in our experiments. When a strength model was formalized and fit to the data from Experiment I, it was found that a two-trace version gave good quantitative fits while a one-trace version did not, suggesting that both short and long-term memory independently contribute in this task.

Although a number of memory theorists (e.g., Kintsch, 1970; Slamecka, 1968; Tulving, 1968; and Tulving and Madigan, 1970) have discussed retrieval processes in recall, an understanding of such processes has proven to be elusive. One approach to this problem is suggested by the study of retrieval processes in recognition memory. That body of research stems from an information processing orientation in which total reaction time (RT) is assumed to be the sum of processing times for additive stages (Sternberg, 1969a). Motivated by these considerations, we have developed a task in which lists of paired associates (PA) are well memorized prior to the experimental session. We measure the time required by S to say the response term following visual presentation of the stimulus term. We conceptualize this time as reflecting three additive stages: 1) encoding the stimulus, 2) retrieving the response, and 3) executing the response.

Our purpose has been to use the paradigm just described to investigate process- oriented models of recall. First, we have provided basic data on the effects of several variables which have proven to be of interest in the study of recognition memory; the sensitivity of RT to the number of items memorized (Sternberg, 1969b), the sequence of probes (Theios, 1973; Theios, Smith,

Haviland, Traupman, and Moy, 1973), and degree of practice (Atkinson and Juola, 1973, in press) should all shed light on the processes involved in recall. Second, having conceptualized the experimental task in terms of three general stages, we have attempted to experimentally isolate the stages; noting the interaction of processing times for each stage with the variables mentioned above, we can draw some conclusions about the nature of the processes involved in each. Specifically, we have been most concerned with establishing that our variables have their effect on the retrieval stage. Third, we have attempted to determine whether semantic memory, as well as episodic memory of the experimental task (see Tulving, 1972), plays a role in this paradigm. If so, the paradigm may prove to be useful in studying the organization of semantic memory.

Prior to reporting our experiments we will discuss several classes of retrieval models we have examined in the context of our recall task. Rather than viewing our experiments as critical tests of strong retrieval models, we have investigated several predictions from a variety of weak models. The substantial data base which we have collected allows the rejection of some classes of models, and suggests which other classes require further specification and testing.

The retrieval processes we will consider may be divided into two classes: serial scanning, which may be exhaustive or self-terminating, and direct access. In recognition memory, Sternberg's (1966) model, which assumes a serial exhaustive

scan of short-term memory (STM) has gained considerable support (see Nickerson, 1972, for a general review). A serial exhaustive scan model may be applied to our PA recall task if we assume that a scan of the list of stimulus members allows access to the appropriate response. The major prediction of this model is a linear increase in RT as a function of list length (RT-L function). In addition, because the scan is assumed to be exhaustive, no variables other than list length (e.g., sequence of probes or serial position of the probed item) should systematically effect RT. If the memorized list exceeds the capacity of STM, and if a match is not found in the initial scan of STM, sampling of items from LTM (with or without replacement) and exhaustive scanning of each sample in STM until a match is found could be hypothesized (see Shiffrin, 1970; and Tzeng, 1972). If sampling is without replacement, it can be shown mathematically that the RT-L function may be non-linear, a prediction which seems intuitively desirable for long lists. Furthermore, if recently presented items (at least the last item) are assumed to have a higher probability of remaining in STM, and thus being scanned in the first sample, these models have the added feature of predicting shorter RT for recently presented items. Note, however, that such sequential effects are a consequence of assuming sampling from LTM and, therefore, would not be predicted for lists shorter than the STM capacity.

An alternative scanning mechanism which has been incorporated into retrieval models of recognition is a self-terminating

process. Three types of self-terminating models can be distinguished. In the first type, order of scan is completely random. Such a model can be falsified by the existence of any systematic within-list difference in RT (e.g., serial position effects) but can not be differentiated from exhaustive scanning models on the basis of recall data. The other two classes of self-terminating models, both of which are distinguishable from exhaustive scan models, share the assumption that the scan always starts at the top of a memory stack. In static models, the order of the stack is the same on all trials and is dictated by serial position or categorical organization of items. In dynamic models, the order of the stack constantly changes, but in a manner determined by the sequence of probes. Failure to find serial position and/or sequential effects can falsify specific formulations of these last two classes of models.

A specific dynamic stack, self-terminating scan model, proposed by Theios (Theios, 1973; Theios, Smith, Haviland, Traupman, and Moy, 1973), has had considerable success in accounting for choice reaction time (CRT) and recognition memory data. The major component of this model is a scanning buffer containing S-R pairs (e.g., in the context of recognition memory, 4-YES, 5-NO). Although Theios and his students have suggested several alternative conceptualizations of the buffer, their most recent tests of the model (Theios and Walters, 1973) support the position that Ss control the size of the buffer and set it equal to the target set size. There are two critical processes in this

model. First, the memory representation of the presented probe has some probability of moving up in the STM stack, and thus pushing down intervening items. Second, a fast serial self-terminating scan of the STM buffer begins when a probe is presented; the process terminates as soon as a response is obtained. Although Theios' model was originally proposed in the context of recognition memory and choice reaction time (CRT) tasks, it is directly extendable to a recall paradigm since there are no restrictions on the S-R mappings; while a many-to-one mapping is used in a recognition task, the one-to-one mapping in a recall task seems to be within the framework of the model. In contrast to exhaustive scanning models, this dynamic stack model predicts that RT for all list lengths decreases with increased probe probability and decreased lag (number of items intervening since the last presentation of the current item). A strong consequence of the assumption that stack size equals the size of the target set is the prediction that RT will vary linearly with the number of PAs in our experiments.

As an alternative to serial scanning of the stimulus set, several investigators (Baddeley and Echob, 1970, 1973; Corballis, 1967; Corballis, Kirby, and Miller, 1972; Corballis and Miller, 1973; Nickerson, 1972; and Wicklegren and Norman, 1969) have hypothesized that RT is a function of the strength of the representation of items in memory. In the context of our PA task, we assume that RT is a monotonic decreasing function of the strength of the association between the probe stimulus and the required

response; "strength" may be the value of a single trace or the number of traces between stimulus and response, (see Howell, 1973, for a discussion of the current state of this distinction). The basic notion is that the strength of an S-R association increases when that stimulus is tested and decreases during the presentation of intervening items. Such a process has several properties. First, increases in RT with increased lag and decreases in RT with increased probe probability are predicted as natural consequences of the incremental and decremental processes. Furthermore, due to the independence of the incremental and decremental processes, a strength model might well be able to simultaneously handle many patterns of lag and repetition effects. The dynamic stack model, which attributes both of these phenomena to the same process, may therefore be more limited in the types of effects it can predict. Second, since the average lag between presentations of an item ordinarily increases with list length, strength will be lower and RT longer for longer lists. Third, strength models predict practice effects (reported for a recognition task by Atkinson and Juola, 1973, in press). Scanning models, on the other hand, can only account for such effects by making the ad-hoc assumption that scan rate increases with practice, a statement which is more descriptive than explanatory. A fourth aspect of strength models is the emphasis they place on LTM. Systematic differences in RT as a function of semantic organization have been reported in recognition experiments by Clifton and Gutschera (1971), Homa (1973),

and Naus, Glucksberg, and Ornstein (1972). If the locus of retrieval is in an organized LTM, then the incorporation of semantic effects seems natural.

To summarize, the discussion of several classes of retrieval models suggests that the shape of the RT-L function over a wide range of list lengths, practice effects, lag and repetition effects, and semantic memory effects are all of interest in ascertaining the nature of retrieval processes in recall. While we have not proposed direct tests of specific retrieval models, the pattern of results from the current experiments shed light on the successes and failings of a wide range of possible models.

Experiment I

The purpose of Experiment I was to evaluate the effectiveness of our paradigm as a means of studying retrieval processes in recall, and to ascertain the effects of list length, practice, lag and repetition on RT.

Method

Design and subjects

Twenty undergraduates at the University of Massachusetts served as Ss. List length was a within S variable, with each S participating in five 45 minute sessions with a different list for each session. List lengths were 3, 6, 9, 12, and 15, and order of list length was Latin squared over Ss (four Latin squares were used). Experimental sessions consisted of eight blocks of 48 trials.

Materials

Each stimulus item was a different category label, and each response item was a frequently given response to that category (Battig and Montague, 1969). Stimuli and responses were between three and six letters long. One third of the stimulus items in each list were 1, 2, or 3 syllables and all words were approximately equated for frequency of use in the English language (Kucera and Francis, 1967).

Procedure and apparatus

A PDP 8/E computer controlled the sequencing and timing of stimuli, and recorded stimuli, responses, and RT for all trials. Stimulus members of the memorized pairs were randomly presented on a video TV monitor (with the restriction that all items were presented equally often), and Ss were required to vocalize the response member as quickly as possible. The S's response triggered a voice key which caused the removal of the stimulus from the screen, and the appearance of the correct response, as well as the S's RT. The S then pulled one of two triggers to indicate whether his response was correct or not. RT was measured from the onset of the stimulus until the triggering of the voice key. The time between consecutive probes was 3.5 seconds and the onset of a probe was preceeded by a .5 second warning tone.

Ss were given experience with a practice list on the day before the first experimental session and ran for one block with the practice list at the beginning of each experimental session. This served to reduce item-independent warm-up effects. Each

day the S received a new list which he took home to memorize in serial order for the following session. Ss were required to give two perfect trials of serial recall before the experimental session began. They were paid a total of $15 for the six sessions and were warned that no pay would be received if they were unprepared for any session.

Results

RT-L function

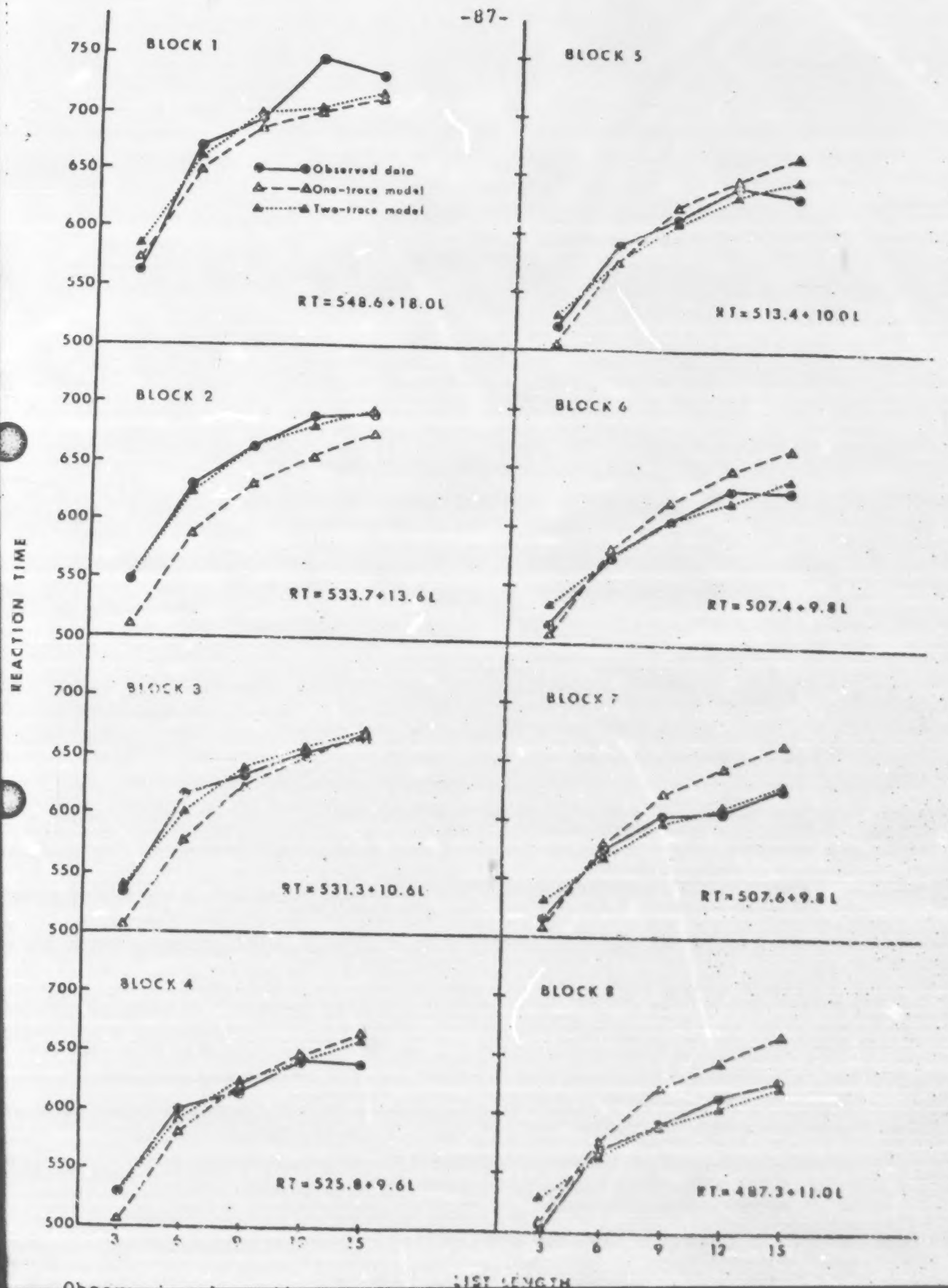
The RT-L functions for each of the eight blocks are plotted in Figure 1 (the solid lines are the relevant observed data). All results are reported in terms of arithmetic means.[2] As can be seen, the curves are all negatively accelerated, a point supported by a trend analysis of the data (see Myers, 1972, Chapter 14) ([$F(1,19) = 9.4$, $p < .01$] for the linear component and [$F(3,57) = 6.8$, $p < .01$] for the deviations from linearity component of the list length effect).

Insert Figure 1 About Here

RT for error trials (machine and S errors) and trials longer than three seconds have been eliminated from all analyses. This accounted for less than six percent of the data in this and subsequent experiments.

Practice effects

As can be seen in Figure 1, RT decreases with practice throughout the eight blocks of the experimental session, but this decline is most marked for early blocks. Trend analysis of


BLOCK 1
Observed data
One-trace model
Two-trace model
RT = 548.6 + 18.0L
BLOCK 5
RT = 513.4 + 10.0L
BLOCK 2
RT = 533.7 + 13.6L
BLOCK 6
RT = 507.4 + 9.8L
BLOCK 3
RT = 531.3 + 10.6L
BLOCK 7
RT = 507.6 + 9.8L
BLOCK 4
RT = 525.8 + 9.6L
BLOCK 8
RT = 487.3 + 11.0L
REACTION TIME
750
700
650
600
550
500
3
6
9
12
15
LIST LENGTH

the block's main effect yields both significant linear [$F(1,19) = 145.5$, $p < .01$], and deviations from linearity [$F(7,114) = 37.3$, $p < .01$] components. A second trend analysis of the list length by blocks interaction reveals that the slope of the RT-L function declines significantly with practice, again most noticably in early trial blocks.

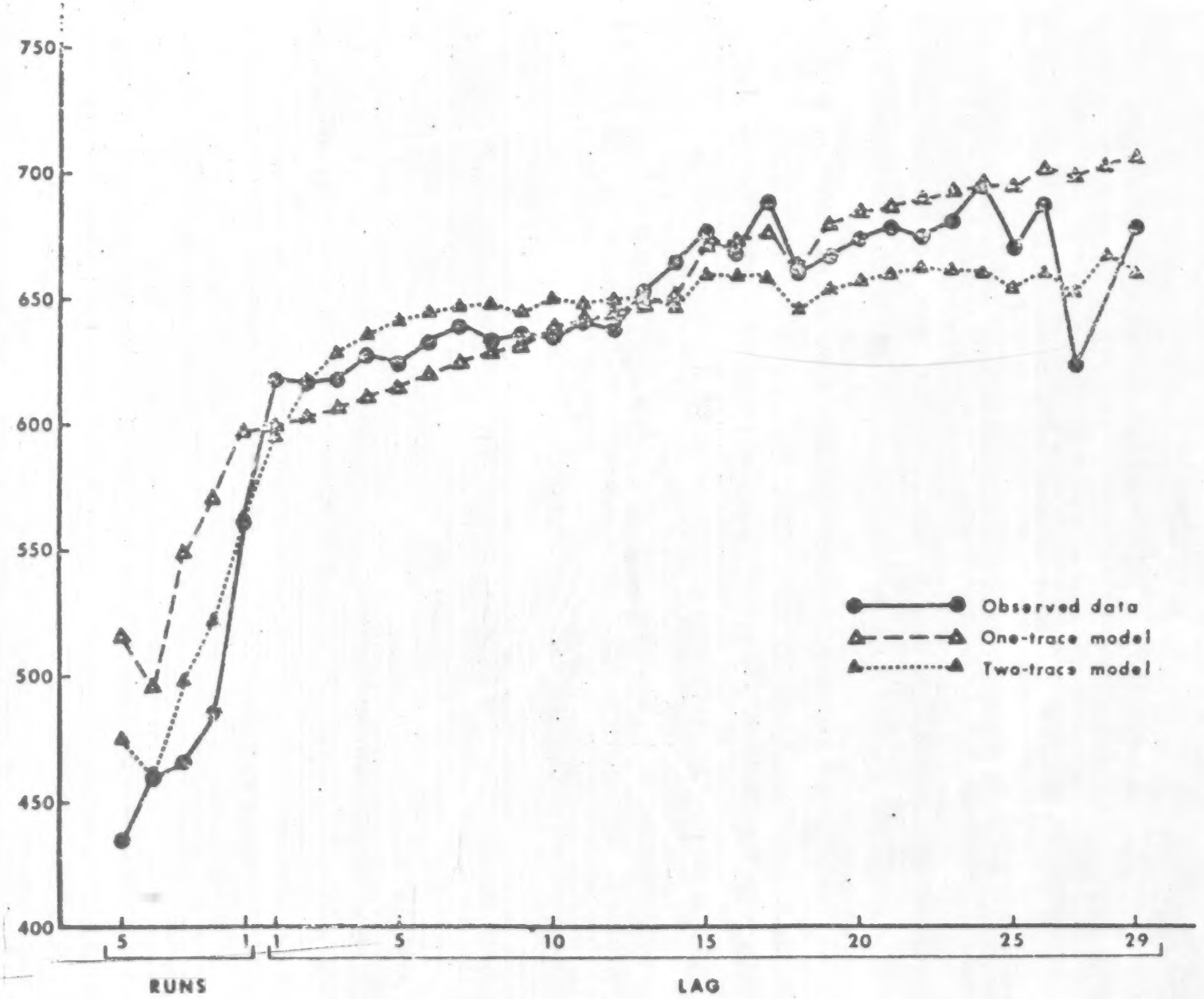
Sequential effects

Lag. RT as a function of lag, averaged over list lengths[3] and blocks is plotted in Figure 2. To assess the slight upward trend in RT as lag increases beyond lag one, as well as any interactions of lag with list length and/or practice, analyses of variance were computed for the slopes of the lag functions.

Insert Figure 2 About Here

A small but reliable upward trend in RT with increase in lag beyond one (mean slope = .57) was observed; the test that the mean slope differs from zero [$F(1,19) = 65.7$, $p < .01$] was significant. This slope is not related to either list length or amount of practice with the list.

Runs. A run is defined as the number of consecutive occurrences of a probe which follow a non-occurrence. As can be seen in Figure 2, RT sharply declines with consecutive repetitions of the same probe. Run data from list length three suggests that RT asymptotes after a number of repetitions (approximately 445 msec. after three repetitions of this list length). Unfortunately there are too few observations of long runs for longer list



Observed and predicted sequential RT in Experiment I

lengths to know whether all list lengths approach the same asymptote. It is clear, however, that run length interacts with list length; the slope of the RT-L function decreases with increased run length (slope = 9.4, 6.4, and 2.9 for runs of length 0, 1, and 2 respectively). The low slope for run length two suggests that list length has little effect on RT with moderately long runs of a single probe.

Other findings

In addition to the list length, practice, lag and repetition effects reported here, serial position effects were examined. End items were responded to most quickly; however, no systematic effects were observed for non-end items, nor for RT as a function of serial distance from the last probe. In addition, number of syllables of the stimulus had a significant effect on RT [$F(2,38) = 5.0$, $p < .05$]; however, the means were not ordered in a systematic manner (mean RT = 618.0, 631.1, and 624.3 msec. for 1, 2 and 3 syllables, respectively). It therefore seems that this difference should be attributed to variability of the specific words chosen rather than to number of syllables.

Discussion

The results of Experiment I point to a number of reliable findings which should be incorporated into a model of recall. To briefly summarize: (1) the RT-L function is negatively accelerated; (2) practice consistently affects both slope and intercept of the RT-L function, but these effects become less marked with increased practice; (3) lag significantly affects

RT for all list lengths; responses to items probed at lag zero are very fast and RT gradually increases with lags beyond one; (4) RT decreases rapidly with repeated presentations of a single probe and this effect is more marked for longer list lengths. These results can be used to narrow the set of viable models previously discussed.

Exhaustive scan models

The entire class of exhaustive scanning models seems untenable. Simple versions predict neither negatively accelerated RT-L functions nor lag and repetition effects; the more complex sampling versions still make the incorrect, strong prediction that lag and repetition effects will not be present when list length is within the capacity of STM. Yet, such effects were observed for list length three. Also, practice effects are not directly predicted by scanning models, but rather must be accounted for by the post hoc assumption that scan rate increases with practice, which implies independence from the specific set of items. If scan rate increases, then the increase should be reflected in lower slopes of the RT-L function over experimental sessions. Again, the data do not support even this indirect account of practice effects; the slope of the RT-L function did not decrease across sessions; mean slopes were 7.7, 6.8, 15.4, 9.3, and 17.8 for days 1 through 5 respectively.

Self-terminating models

As we have already pointed out, self-terminating models can make a wide variety of predictions, and thus specific instances,

rather than the entire class must be falsified. The absence of serial position effects, and the presence of sequential effects suggest that any model which posits scanning on the basis of serial position, serial distance from the past probe, or random scanning, is inappropriate. On the other hand, Theios' dynamic stack model does predict effects of the sequence of probes. Both lag and repetition effects follow from the stochastic reordering of the buffer on each trial. Whether this single mechanism would permit a good fit to both the small lag effects and large repetition effects observed in our experiment is not clear. Two other possible problems exist. First, the currently accepted version of the stack model (Theios and Walters, 1973) assumes a buffer equal in size to list length. This leads to the prediction of linear RT-L functions which we do not obtain. Since it is not clear what, if any, limitation on buffer capacity Theios and Walters envisage, the model currently fails to capture a significant aspect of the data. Second, like the exhaustive scanning models, the stack model lacks an integral mechanism which predicts practice effects. In addition, it is not clear how the model would be extended to encompass effects of associative strength (or other organizational properties of LTM) which we will consider in subsequent experiments. None of these considerations demonstrate that the stack model is wrong in its fundamental assumptions but rather that it is inadequately developed with respect to the role of LTM to account for certain of our findings - effects of long lists, practice, and semantic variables.

Strength models

Instead of considering a general class of strength models, we will at this point consider one specific formalization of a strength model. A number of assumptions are possible with respect to how strength changes over trials and how RT is mapped from strength. Our particular choice of assumptions has been directed largely by considerations of mathematical tractability; however, the model appears capable of providing a qualitative account of the basic results of this experiment.

We assume a strength continuum, s, such that:

$$0 \leq s_{min} \leq s \leq 1 \qquad (1)$$

where s_{min} is a lower bound on s corresponding to some minimum level of strength and related to semantic characteristics of the item; for example, s_{min} might be higher for S-R pairs that are more strongly associated. Further, we assume that retrieval time is inversely proportional to the distance between current and maximum strength[4]; therefore, total RT is

$$T = W + k(1-s) \qquad (2)$$

where W is assumed to be a constant time for encoding the stimulus and executing the response (support for this assumption is obtained in subsequent experiments). In addition, we assume that the incremental and decremental operators are linear functions of current strength. When an item is presented on trial $n-1$:

$$s_n = s_{n-1} + \alpha_1(1 - s_{n-1}) \qquad (3)$$

and if the item is not presented on trial $n-1$:

$$s_n = s_{n-1} - \alpha_2(s_{n-1} - s_{min}) \qquad (4)$$

$$0 \leq \alpha_1, \alpha_2 \leq 1$$

Negatively accelerated increases in the RT-L function for trial 1 data suggested an additional assumption about $s_1(L)$, strength on trial 1 for a list of length L.

$$s_1(L) = s_{min} + \delta^{L-1}(1 - s_{min}) \qquad (5)$$

$$0 \leq \delta \leq 1$$

Equation (5) has the properties that: 1) $s_1(L)$ approaches s_{min} as L increases, yielding an increasing gradient of trial one RT with increases in list length, and 2) $s_1(1) = 1$, which seems plausible.

Derivations from this model are presented in Appendix I. The predicted RT-L function will clearly be increasing and negatively accelerated as observed. We also noted earlier that mean RT decreases over trials. Our derivations indicate that under appropriate parameter values the specific model under consideration can account for the stronger result obtained in our experiment, the decline in slope of the RT-L function over trial blocks. The model also predicts, independent of parameter values, that lag functions will increase in a negatively accelerated manner and slopes and intercepts of the RT-L function will decrease as number of consecutive repetitions increases. These results were all observed.

Encouraged by the qualitative capability of our model, we estimated parameters and fit RT-L functions for each trial block.

The estimate of W, the sum of encoding and execution time, is based on time to read aloud the probe word; such data are obtained in subsequent experiments. The parameters, k, s_{min}, α_1, α_2, and δ were estimated from the mean RT for each list length for each block using STEPIT (Chandler, 1966) a minimization subroutine, to achieve a least-squares solution.[5] The predicted RT-L functions did show negative acceleration and the slopes and intercepts of these functions did decrease with practice. However, the model failed in two respects. While the predicted RT-L functions decrease in slope from block two on, the size of this effect was less marked than in the observed data and it was not present for block one. Perhaps even more serious is that when we inserted these parameter estimates into expressions for lag and repetition functions (see Appendix I), the fits were extremely poor. Parameters which allow sufficient practice effects do not predict sizeable lag and repetition effects although, again, the model correctly predicts the direction of these effects, and the fact that the repetition effects are larger than the lag effects.

Strength models with alternative RT mappings could be developed. However, we believe that any model which relies on a single linearly increasing and decreasing strength process will encounter the same difficulties noted here. In order to adequately handle the sizeable and prolonged practice effects exhibited in the data, a small incremental parameter is required. But, as demonstrated in the derivations of Appendix I, predic-

tions of a decrease in the slope of the RT-L function with practice requires the decremental parameter to be even smaller. This in turn will incorrectly lend to extremely small sequential effects.[6]

One way to deal with the problems just noted would be to retain our current strength model to handle long-term effects, and to add a short-term mechanism to deal with the sizeable sequential effects. Such a mechanism might be Theios' short-term stack, a one-slot buffer holding the last item, or a short-term trace (Wicklegren, 1970, 1972) with large incremental and decremental parameters. We have fit a model of the last type, employing the assumption that total strength is the sum of a short-term (σ) and a long-term (λ) trace (see Wicklegren, 1970). Thus,

$$T = W + K(1 - \frac{\sigma + \lambda}{2}) \qquad (6)$$

Separate incremental and decremental parameters were estimated for the two memory traces. In addition, the s_{min} and δ associated with the long-term trace were estimated. Assuming that the short-term process reflects a limited capacity system, we set minimum short-term strength at zero and initial short-term strength at $1/L$. In order to simultaneously estimate short and long-term parameters of this model, both practice and sequential data were used; the criterion of best fit was the minimum root mean squared deviation (RMSD) of observed from predicted RT for all lag and run lengths for each of the eight blocks, weighted by number of observations for each point.

For the purpose of comparison, this procedure was also used to estimate parameters and fit the one-trace model. The parameter estimates and RMSD for both models are presented in Table 1. As would be expected, the two-trace model (with two

---

Insert Table 1 About Here

---

additional parameters) gives a better fit than the one-trace model does. Several aspects of the parameter estimates are consistent with expectations from the literature on short and long-term memory; the incremental and decremental parameters are considerably larger for the short-term trace than for the long-term trace, and in fact, the decremental process for the long-term trace is essentially nonexistent ($\alpha_2 < .001$).

Observed and predicted RT-L functions for each block are presented in Figure 1, and observed and predicted sequential data averaged over blocks and list lengths, are presented in Figure 2. From Figure 1 we can see that while the two-trace model slightly over-predicts the decrease in slope of the RT-L function with practice, the one-trace model under-predicts this effect to a considerably larger degree. In addition, Figure 2 indicates that the one-trace model does not predict a sufficiently large effect of run length, while at the same time it predicts too large a lag effect. On the other hand, the two-trace model fits the sequential data quite well. It thus appears that with the addition of a second trace, at least one formalization of a strength model can adequately capture both the qualita-

Table 1

Parameter estimates of strength models*

	One-trace model	Two trace model	
RT_n =	$W + k(1 - S_n)$	$W + k(1 - \frac{\sigma_n + \lambda_n}{2})$	
W* =	425	425	
K =	756.6	332.2	
RMSD =	45.4	35.6	
		σ(short-term trace)	λ(long-term trace)
S_{min}	.500	assumed = 0	.000
α_1	.151	.464	.059
α_2	.022	.440	.0008
δ**	.297	----	.670

*W = 425 is estimated from the READ data of Experiments II-IV. In addition, $\sigma_1(L) = 1/L$ by assumption; therefore, no δ is estimated for the short-term trace.

tive and quantitative results of Experiment I.

To summarize, we have shown that simple scanning models do not adequately predict even the qualitative aspects of the data from Experiment I. Strength models can be developed which make the correct qualitative predictions, but several shortcomings were noted when one such one-trace model was fit to data. A two-trace strength model was developed which was able to capture the quantitative aspects of the data.

Experiment II

Experiment II was designed to ascertain whether the RT-L, practice, and sequential effects demonstrated in Experiment I could be attributed to the encoding stage, or were solely accounted for by the retrieval stage. This was accomplished by alternating blocks where the S was required to recall the response item to the visually presented stimulus (RECALL blocks) with blocks where the S was required to read aloud the stimulus (READ blocks). Presumably, the second task involves only the encoding and execution stages whereas the first task also involved retrieval.

Smith (1968) has reviewed several experiments dealing with the effects of list length upon encoding time. For highly compatible stimuli and responses, list length had either little or no effect. In the study most closely related to ours, Pierce and Karlin (1957), using a continuous reading task, found that variation in vocabulary size from 2 to 256 words had little effect on rate of reading random words equated for length and familiarity.

While this result suggests that list length effects are primarily attributable to retrieval, we wish to verify this finding for our paradigm. Practice and sequential effects in the reading task are also of interest.

In addition, in the present experiment, we manipulated associative strength (AS) of the PAs. AS was defined by using response items which were instances (associated PAs) or non-instances (non-associated PAs) of the stimulus category. Several concerns motivated this manipulation. First, systematic effects of AS on RT would suggest that our recall paradigm may be of some use in studying the organization of LTM. Second, such a manipulation provides a further qualitative test of several models. Serial scan models do not predict effects of AS. On the other hand, strength models, which place an emphasis on the role of LTM in retrieval, could incorporate AS effects. For example, the strength model developed in the context of Experiment I makes strong predictions if we assume s_{min} (the asymptotic minimum strength an item can decay to) increases with AS. Derivations in Appendix I demonstrate that RT should be shorter for associated pairs, and that the slope of the RT-L and lag functions should be lower for associated pairs.

Method

Design and subjects

Six undergraduates at the University of Massachusetts served as Ss. List length was a within-S variable, with each S participating in three 45 minute sessions. List lengths were 6, 12,

and 18 PAs and order of lists was counterbalanced over Ss. Half of each list was associated PAs. Each experimental session consisted of eight blocks of 48 trials each. Blocks alternated between READ and RECALL blocks, with half of the S having each type of block for block one.

Materials

Each stimulus item was a different category label and half of the response items in each list were frequently given responses to that category (associated PAs): the other half were unrelated nouns (non-associated PAs). Stimuli and responses were between three and six letters. One third of the stimulus items in each list were 1, 2, or 3 syllables and all words were approximately equated for frequency of use in the English language (Kucera and Francis, 1967).

Procedure and apparatus

The procedure and apparatus were the same as in Experiment I with the exception that half of the blocks were READ blocks.

Results

Basic Findings

The baseline findings from Experiment I were replicated in Experiment II; 1) RECALL RT was a negatively accelerated function of list length, 2) the slope and intercept of the RECALL RT-L functions decreased with practice, 3) RECALL RT increased with lag and decreased with runs of the same probe, and 4) serial position did not systematically effect RECALL RT.

Figure 3 presents the RT-L functions for READ and RECALL

blocks plotted separately for associated and non-associated PAs. The gross aspect of this figure, flat READ functions which are not affected by AS, and increasing RECALL functions which do differ with AS, meet our general expectations. We will now examine them more closely.

Insert Figure 3 About Here

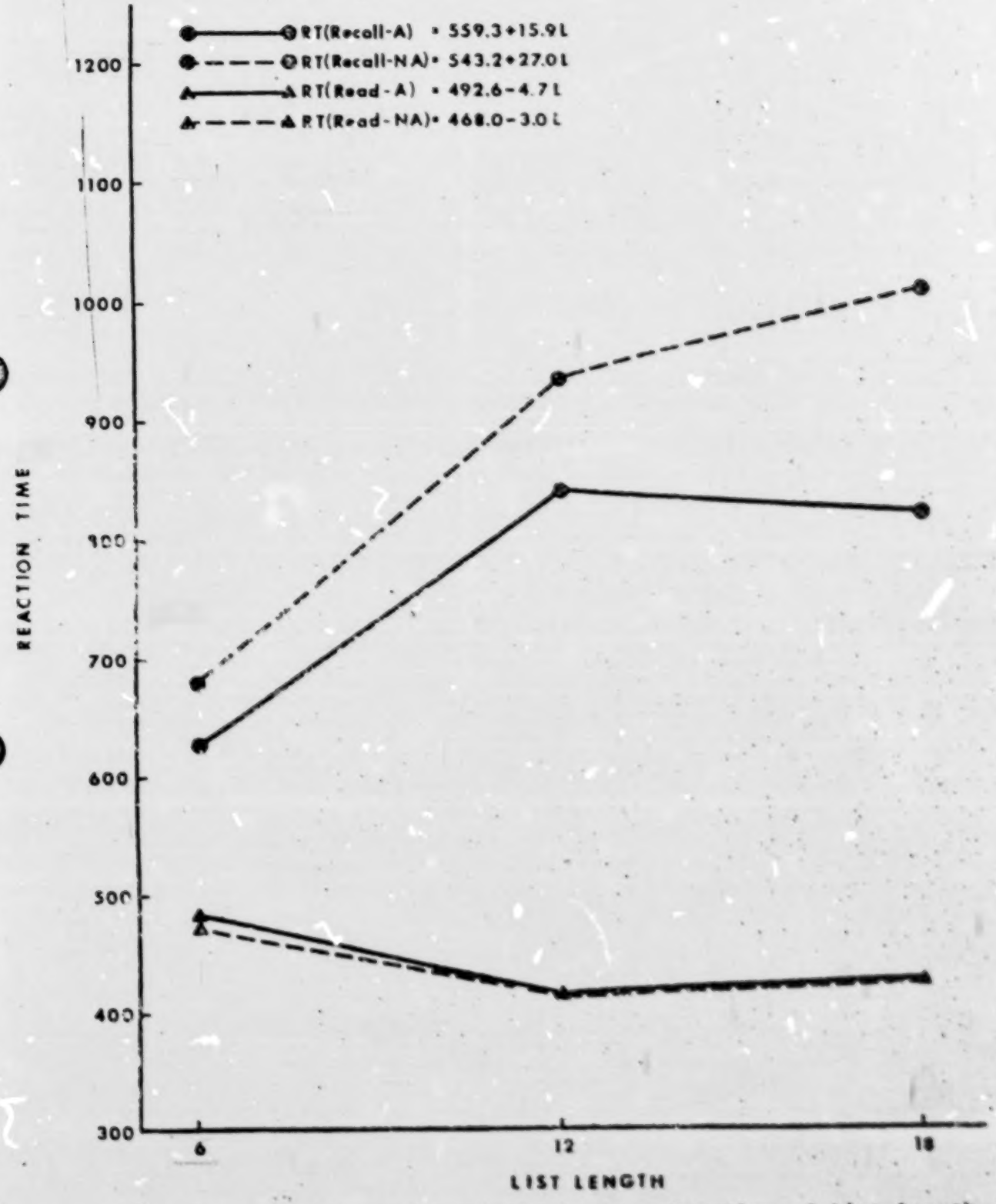
READ task

READ time was not affected by list length, practice, nor AS. This finding is especially impressive when one notes the considerable amount of power to detect such effects due to the extremely small error variance in the READ data. While the lag functions, which can be seen in Figure 4, appear to be flat for the READ data, the slight upward trend (slope = .54) was reliable [$F(1,4) = 11.6$, $p < .05$].

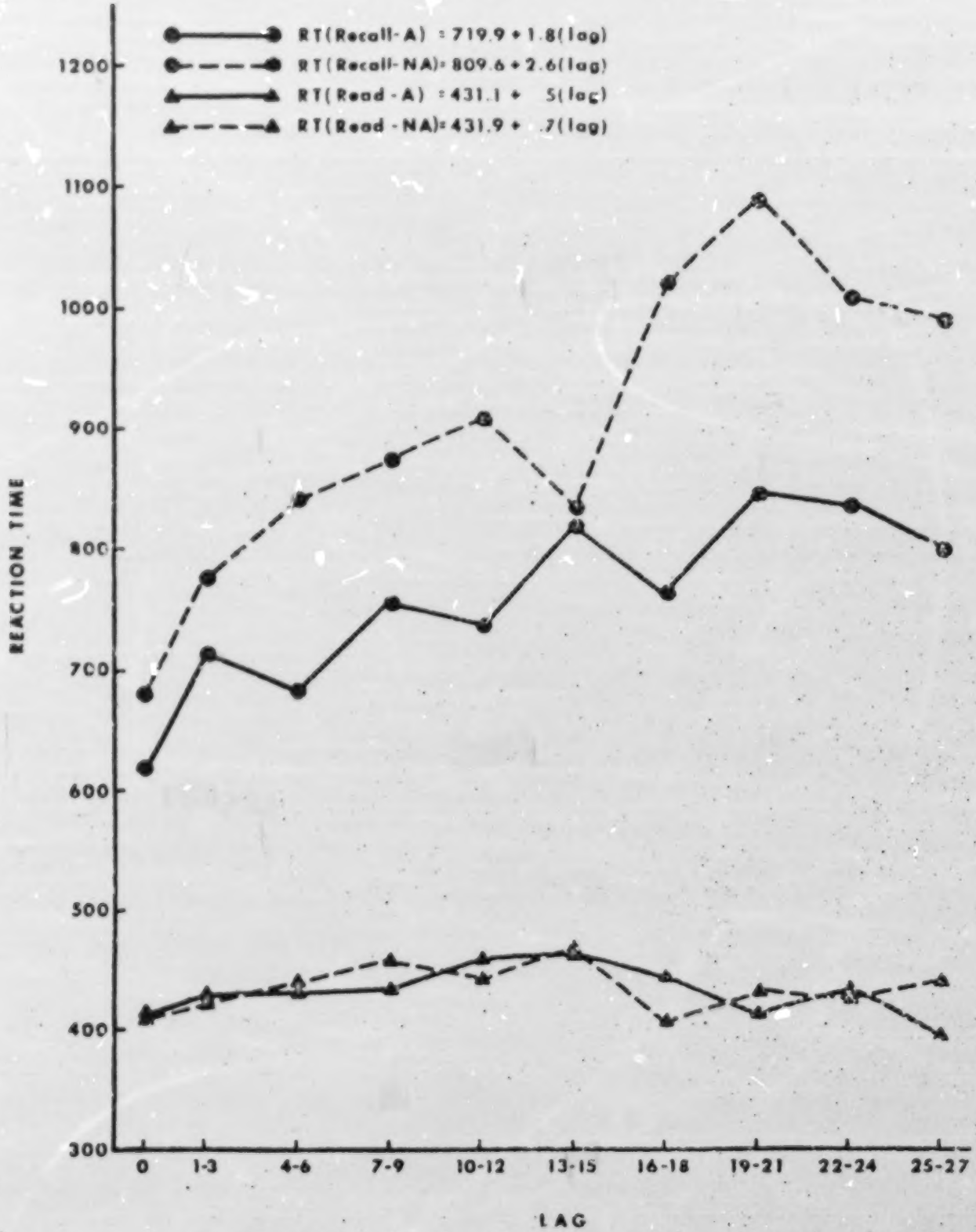
Insert Figure 4 About Here

Associative strength of PAs

Mean RECALL RT is lower for associated pairs. The marginal significance of this effect [$F(1,4) = 7.3$, $.05 < p < .1$], appears to attest more to the lack of power resulting from only four error df than to a lack of reliability; the effect was observed in all Ss and is significant in two subsequent experiments employing other samples of PAs. The interaction of AS with list length is significant [$F(2,8) = 8.5$, $p < .05$]. A trend analysis supports the conclusion that the slope of the RT-L function for



Mean RT for READ and RECALL blocks as a function of list length and AS in Experiment II



READ and RECALL lag functions in Experiment II

non-associated pairs (27.0) is steeper than the slope for associated pairs (15.9), [$F(1,5) = 13.1$, $p < .01$].

In addition consistent with the predictions of the strength model (see Appendix I) practice and lag effects were more marked for non-associated PAs than for associated PAs, although these findings failed to reach statistical significance.

Discussion

Experiment II served to replicate the basic findings of Experiment I using different materials. In addition, a major question of interest in Experiment II was whether the variability in RT could be attributed solely to encoding processes. The READ time, assumed to reflect encoding and execution stages, was not significantly affected by list length or practice, and lag contributed very slightly in this task. Thus, the identification of these effects with the retrieval stage seems to be well founded. Furthermore, using the READ data we can estimate W, the encoding and execution time parameter of the strength model, to be 425 milliseconds.

The effect of AS on the retrieval process was also of interest in the present experiment. As indicated, not all of these effects are significant in our data. Establishment of their reliability is of some interest because they provide interpretive difficulties for serial scanning models which, without explicitly incorporating assumptions about the role of LTM, predict neither main nor interaction effects involving AS. Furthermore, the fact that the trends of the main and interaction

effects of AS are all in the direction predicted by the strength model (see Appendix I) suggests the validity of the strength concept as well as the identification of minimum level of strength with AS. It is also interesting to note that an analog to the AS manipulation in CRT experiments, (see Smith, 1968, for a general review), S-R compatibility, has also yielded results consistent with the predictions of the strength model: Conrad (1962), using nonsense syllables and words as stimuli, found that the effect of both practice and number of choices is more marked for low-compatible material, and Kornblum (1973) in a review of the CRT literature, has noted that sequential effects are greater with low-compatible material.

Experiment III

The results of Experiment II are not entirely conclusive with respect to the role of the AS variable. In Experiment III, we again obtained RT for associated and non-associated pairs; however, this time AS was a between, rather than a within-list, variable. We are again interested in determining whether slopes of RT-L and lag functions are steeper, and practice effects are greater, for non-associates; recall that both the strength model and the analogy with manipulations of S-R compatibility in CRT experiments predict such results.

An important distinction between models we have considered is that scanning models attribute list length effects to the number of items being scanned, while strength models attribute these effects to probe probability. Typically, these two mani-

pulations have been confounded (probe probability = 1/L): in this experiment we have varied relative frequency of items within a list as well as list length. Hence, finding faster RT for items that are presented more frequently within a list, would lend support to the strength conceptualization. Theios' dynamic stack model makes similar predictions since more frequent items would be expected to be found at higher positions in the memory stack. A prediction of the strength model that Theios' model does not make, is that relative frequency effects should be more marked for non-associates than for associates.

A basic assumption of our strength model is that the unit which gains or loses strength is the S-R association. In this experiment, we attempted to verify this by including probes with two-to-one S-R mappings; that is, some responses are paired with two different stimuli. Such responses should be no faster than those paired with a single stimulus if frequency of the S-R unit, as opposed to response frequency, is the critical determiner of RT. Furthermore, although RT should increase with lag from the last occurrence of the S-R unit, it should not increase with lag from the last occurrence of the response, if the response was paired with a different stimulus.

If either or both of these predictions are not confirmed, there are two possible interpretations. Our assumption that the S-R unit is the locus of changes in strength may be incorrect; strength of the response term may independently influence retrieval time. The alternative explanation is that there are frequency

and/or lag effects associated with the response execution stage. Absence of frequency and recency effects in the READ data, which presumably includes execution time would support the first hypothesis. Given these considerations, it appeared desirable to again collect READ data, this time in the context of the present design.

To summarize, the points of major interest in Experiment III were: 1) to further assess the main and interaction effects of AS, 2) to examine relative frequency effects when manipulated within a list for both READ and RECALL tasks, 3) to examine the role of response frequency and recency, and 4) to replicate and extend to longer lists, the READ findings from Experiment II.

Method

Design and subjects

Six undergraduates at the University of Massachusetts served as Ss. List lengths were 16, 20, and 24. Each S ran for six experimental sessions which differed with respect to combinations of list length (16, 20, 24) and AS (associated and non-associated PAs). Order of lists were blocked by AS with half of the Ss receiving each level first; order of list length for each level of AS was counterbalanced over Ss. Each experimental session consisted of eight blocks of 50 trials each. Blocks alternated between READ and RECALL blocks, with half of the S having each type of block for block one.

Materials

A pool of one syllable nouns of three to six letters in

length which fell in the A or AA categories in the Thorndike Lorge (1944) frequency of occurrence in the English language norms was established. Pairs of seemingly associated words (rated by the experimenter) were arbitrarily assigned as stimulus or response members of the associated lists. The non-associated lists were constructed by randomly sampling (without replacement) words from the remaining pool.

Procedure and apparatus

The procedure and apparatus were the same as in Experiment II except for the assignment of probes. One half of the items had a one-to-one S-R mapping; as in our previous experiments, each stimulus had a different response associated with it. Relative frequency of these items was varied; frequent PAs (FF) appeared twice as often as infrequent PAs (II). The remaining half of the items (IF) had two-to-one S-R mappings (pairs of different stimuli had the same word as a response). As the mnemonic IF indicates, the stimulus member appeared as often as in the II items and the response was made as often as in the FF items. To clarify this, consider a representative sequence of items for an associated four item list:

FOOT - HEEL	(FF)
FOOT - HEEL	(FF)
NEST - BIRD	(IF)
WING - BIRD	(IF)
HEAD - HAT	(II)

Results

Basic findings

The results of Experiment III are generally consistent with the baseline results (practice, lag, and repetitions) established in Experiments I and II. The RT-L functions presented in Figure 5, however, differ somewhat in shape from those of previous experiments. This is especially true of the difficult-to-learn and more variable list of non-associated PAs.

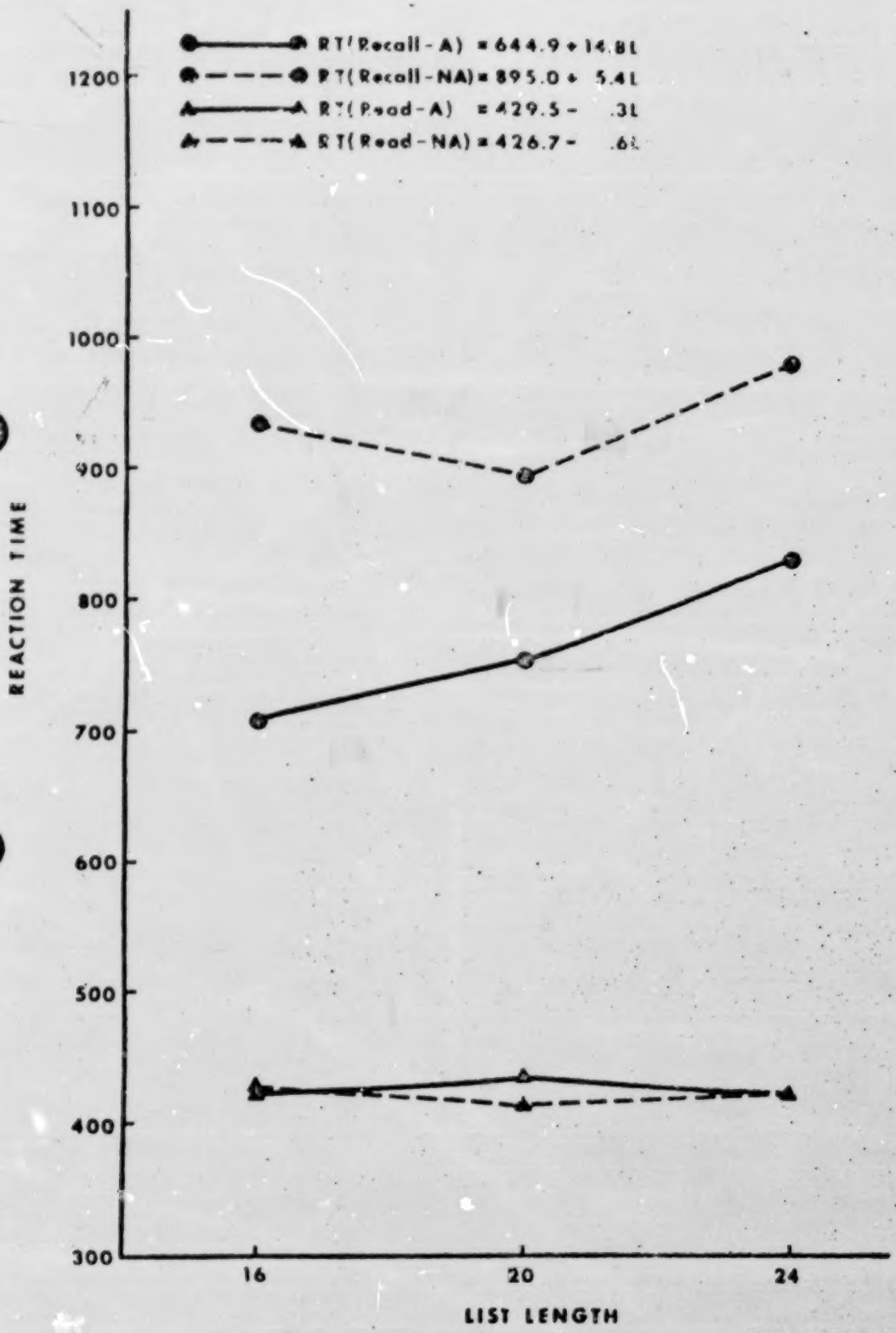
Insert Figure 5 About Here

Associative strength of PAs

As indicated by the RT-L functions of Figure 5, RT is consistently faster to associated PAs [$F(1,4) = 8.4$, $p < .01$], which suggests that this trend in Experiment II was a reliable result. The strength model predicts more pronounced effects of list length and probe probability for non-associates than for associates. The AS by list length interaction, which was observed in Experiment II, was not replicated. However, within each list, there is a trend in the predicted direction for AS to interact with probe probability (mean RT to the relevant FF and II probes are presented in Table 2).

Insert Table 2 About Here

In addition, consistent with the predictions of the strength model, and the trends in Experiment II, the practice and lag effects are more marked for non-associated PAs than for assoc-



Mean RT for READ and RECALL blocks as a function of list length and AS in Experiment III

Table 2

Stimulus and response probability effects

(Experiment III)

		Associates				Non-Associates				Mean
		16	20	24	Mean	16	20	24	Mean	
II	Recall	761.1	780.6	854.6	798.8	1084.3	895.4	1074.4	1018.0	908.4
	Read	414.9	426.7	431.3	424.3	427.3	436.3	415.5	426.4	425.4
IF	Recall	689.1	799.9	836.6	775.3	932.2	999.0	1021.6	984.3	879.8
	Read	433.9	434.4	413.9	427.4	416.9	415.7	427.9	420.1	423.8
FF	Recall	708.3	695.6	806.2	736.7	873.7	812.8	890.6	859.0	797.9
	Read	419.2	441.0	423.1	427.8	441.1	400.7	423.9	421.9	424.9

iated PAs although these effects again fail to reach statistical significance.

Relative frequency effects

In this experiment, FF probes were presented twice as often as II probes, and thus we can assess the relative frequency effect when this variable is manipulated within a list. As can be seen from Table 2, the FF items were responded to more quickly (mean RT = 797.9 msec.) than the II items (mean RT = 908.4 msec.) in the RECALL task. The statistical significance of this finding is attested to by the significant Bonferroni *t* test [$t(5) = 4.2$, *EW* < .05]. No such difference was found for the READ task (mean RT = 424.9 msec. and 425.4 msec. for the FF and II probes respectively).

Response effects

In previous experiments, repetition of a response was confounded with repetition of the S-R association. Inclusion of probes with two-to-one S-R mappings (IF probes) in Experiment III allows us to isolate the effect of repetition of the response term.

Relative frequency effects. Refer back to mean RT to II, IF and FF probes presented in Table 2. Since the probability of the stimulus term was the same for the II and IF probes, the significant difference in RT to II and FF probes (mean difference = 28.6 msec.), should be attributed to the contribution of repetition of the response term, [Bonferroni $t(5) = 3.9$, *EW* < .05]. However, response repetition is not solely responsible

for variation in RT as evidenced by a significant difference in RT (mean difference = 81.9 msec.) between FF and IF probes, [Bonferroni $t(5) = 4.9$, $EW < .05$].

Sequential effects. Just as we assessed the effect of response frequency, we can evaluate the effect of response recency by considering lag functions. These are presented separately for lists of associated and non-associated PAs in Figure 6. Each point on the curves is based on an unweighted average of mean RT for the three list lengths.

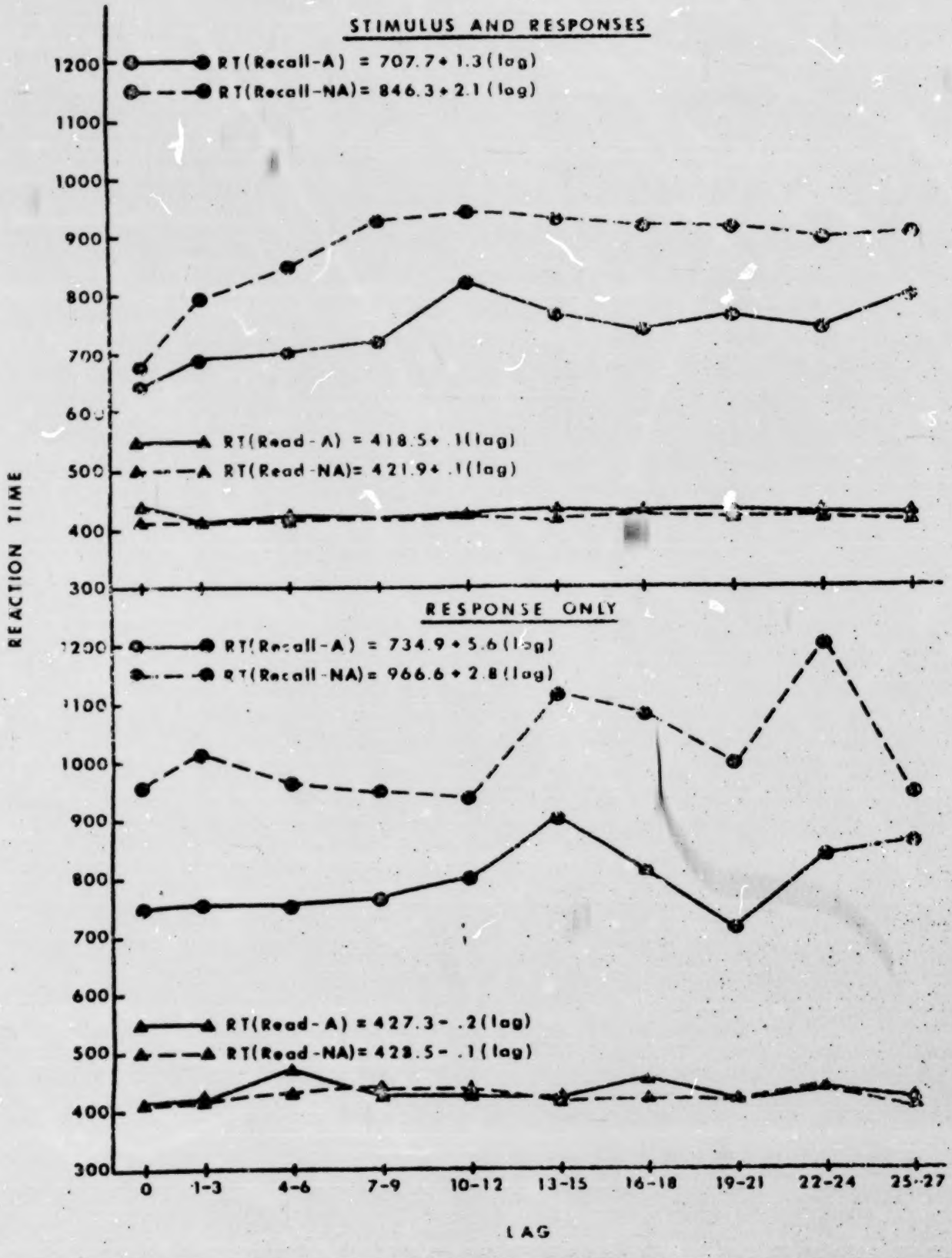
Insert Figure 6 About Here

In the response-only repetition curves (lower panel), lag was measured from the last occurrence of the response term, provided that it had been paired with a different stimulus. While the slope of 4.2 does not differ significantly from zero, it suggests the presence of response specific processes.

In the stimulus and response repetition curves (upper panel) as in the previous experiments, lag was measured from the last repetition of the stimulus and response terms, and hence reflects both stimulus and response processes, as well as processes associated with the S-R association. The slope of this function (mean slope = 1.7) was significantly greater than zero [$F(1,4) = 85.5$, $p < .01$], and is consistent with the lag effects observed in Experiments I and II.

READ task

The findings from the READ task in Experiment II were repli-


STIMULUS AND RESPONSES
RT(Recall-A) = 707.7+1.3(lag)
RT(Recall-NA)= 846.3+2.1(lag)
RT(Read-A) = 418.5+.1(lag)
RT(Read-NA)= 421.9+.1(lag)
RESPONSE ONLY
RT(Recall-A) = 734.9+5.6(lag)
RT(Recall-NA)= 966.6+2.8(lag)
RT(Read-A) = 427.3-.2(lag)
RT(Read-NA)= 428.5-.1(lag)
REACTION TIME
1200
1100
1000
900
800
700
600
500
400
300
0
1-3
4-6
7-9
10-12
13-15
16-18
19-21
22-24
25-27
LAG

cated in Experiment III using longer lists; READ time was not affected by list length, practice, or AS. While there was a significant positive slope to the lag function [$F(1,4) = 8.0$, $p < .05$] this slope was extremely small (mean slope = .12), and no lag zero effect was observed. Furthermore, as mentioned above, when probe frequency was manipulated within a list, READ time was not affected.

Discussion

The effects involving the semantic variable are consistent with the trends observed in Experiment II, predictions of the strength model, and results of S-R compatibility manipulations in CRT experiments. Practice, sequential and relative frequency effects are all more marked for non-associates. The single exception is the lack of slope differences for the RT-L functions of associated and non-associated lists. The replicability of these findings with different materials is encouraging, and presumably the fact that the interactions often failed to reach statistical significance can be attributed to the small number of error df available for most of the significance tests.

One objective of Experiment III was to investigate the effect of varying probe probability when manipulated within a list. Consistent with the predictions of the strength model, RT decreased as probe probability increased. Such effects were not observed in the READ data, a result consistent with our assumption that the locus of the probe probability effect is in the retrieval stage. Experiment III provides firm support for the assumption

that our variables have there effects on the retrieval stage of processing. Using lists longer than previously employed, we again found no effects of probe probability, practice, or lag zero on READ time.

Because the READ data clearly demonstrate no systematic effects in the execution stage, and the response-only lag function was not unequivocally demonstrated to differ from zero, we will postpone discussion of the isolation of response processes until further evidence is presented from Experiment IV.

Experiment IV

Significant effects of response frequency were demonstrated in Experiment III; however, the status of response recency effects, as evidenced by lag statistics, was less clear. Because only four blocks of RECALL data were collected and long list lengths were used (16, 20, and 24) less than five percent of the data contributed to the response-only lag functions. Therefore, in order to provide a more powerful test of the response recency effect, shorter list lengths (4, 8, and 12) were used in Experiment IV (PAs were a subset of those in Experiment III), and eight blocks of RECALL data were collected, followed by a single block of READ data. Furthermore, while main and interaction effects of AS were replicated in two experiments, we hoped to either demonstrate them significantly in Experiment IV, or at least replicate them a third time.

With the exception of the above mentioned changes (list lengths and number of RECALL and READ blocks) and the fact that

twelve Ss were run, Experiment IV employed the same procedure as Experiment III.

Results

Basic findings

The RT-L functions for Experiment IV are presented in Figure 7. The RECALL curves are again increasing functions of list

---

Insert Figure 7 About Here

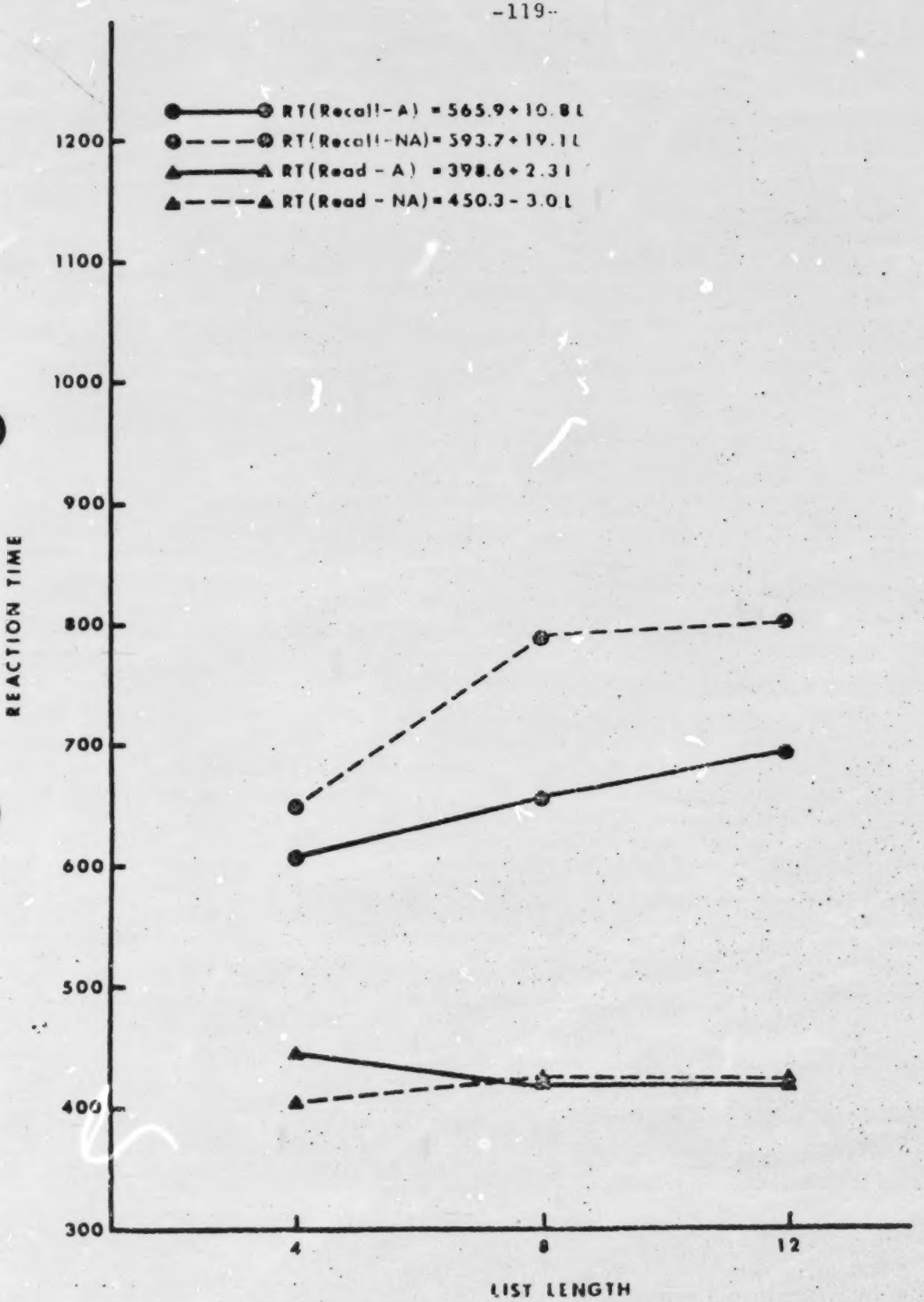
---

length, and RT is faster to associated PAs than to non-associated PAs. The READ curves are again flat, and independent of list length and AS. Both the slopes and intercepts of the RECALL functions decrease with practice, while the READ curves are not affected by practice. Sequential effects are also consistent with those obtained in previous experiments, and will be discussed more fully below.

Associative strength of PAs

Consistent with previous findings, RT is consistently shorter to lists of associated PAs [$F(1,10) = 44.5$, $p < .01$]. As in Experiment II, the slope of the RT-L function is steeper for non-associated PAs (mean slope = 19.1) than for associated PAs (mean slope = 10.8),([$F(1,11) = 5.7$, $p < .05$] for the linear trend component of the AS by L interaction). The manipulation of relative frequency within lists (relevant means are presented in Table 3) also indicates a more marked frequency effect for non-associated PAs [$F(2,20) = 6.4$, $p < .05$].

Insert Table 3 About Here



Mean RT for READ and RECALL blocks as a function of list length and AS in Experiment IV.

Table 3

Stimulus and response probability effects

(Experiment IV)

		Associates				Non-Associates				Mean
		4	8	12	Mean	4	8	12	Mean	
II	Recall	609.2	704.5	700.5	671.4	609.3	844.6	856.0	770.0	720.7
	Read	444.2	406.6	424.9	425.2	420.6	425.0	421.3	422.3	423.8
IF	Recall	583.8	636.6	694.3	638.2	687.4	828.5	761.8	759.2	698.7
	Read	444.7	420.9	422.0	429.2	406.5	436.8	416.9	420.1	424.7
FF	Recall	629.2	653.9	687.7	656.9	628.6	725.3	813.1	722.3	689.6
	Read	441.9	417.5	415.3	424.9	393.0	413.9	429.5	412.1	418.5

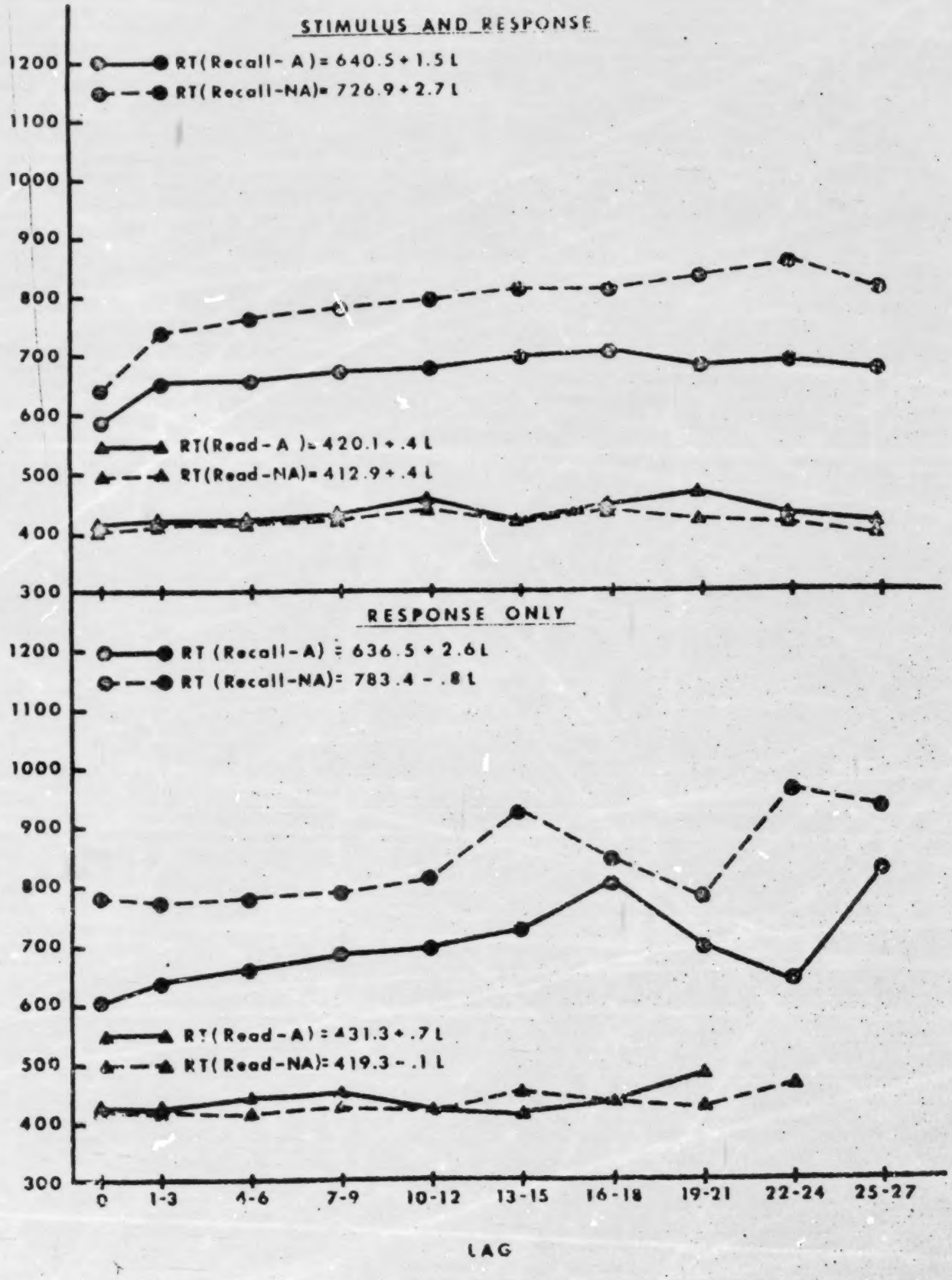
In addition, as in Experiment II and III, practice and lag effects are more pronounced for non-associated PAs than for associated PAs. In this experiment both effects reached statistical significance; [$F(7,70) = 6.4$, $p < .01$] and [$F(1,10) = 29.9$, $p < .01$] for the AS interactions with practice and lag respectively.

Response effects

Relative frequency effects. Mean RT to II, IF, and FF probes are presented in Table 3. As in Experiment III, the mean RT to the three probe types are ordered in the predicted direction (mean RT = 720.7, 698.7, and 689.6 msec. to II, IF, and FF probes respectively). Specifically, the difference in RT to II and IF items [Bonferroni ($t(11) = 2.3$, $EW < .1$], indicates the presence of response specific components of the relative frequency effects.

Sequential effects. The lag functions for Experiment IV are presented separately for lists of associated and non-associated PAs in Figure 8. Each point on the curves is based on an unweighted mean of RT for the three list lengths. A significant increase in RT with response-only lag was observed (mean slope = 1.7; [$F1,10) = 9.6$, $p < .025$]. In the stimulus and response repetition lag function, a reliable slope was also obtained (mean slope = 2.1; [$F(1,10) = 87.1$, $p < .01$]).

Insert Figure 8 About Here



READ and RECALL lag functions in Experiment IV

Discussion

A systematic pattern of findings is obtained when AS is manipulated in this paradigm. Although interaction of this variable with list length, practice, lag, and relative frequency sometimes failed to reach statistical significance, they were replicated in three experiments under different conditions, and all reached conventional significance levels in Experiment IV. In Experiment II, AS was manipulated within lists with category labels as stimuli and category instances or non-instances as responses, while in Experiments III and IV, AS was a between-list manipulation and PAs were either associated or non-associated noun pairs. The direction of the AS interactions conforms with predictions derived from the strength model by identifying s_{min} with AS (see derivations in Appendix I), and with effects of S-R compatibility in CRT experiments (e.g., Hawkins, MacKay, Holley, Friedin, and Cohen, 1973).

The replicable result that frequency and recency of the response member influences RT leads us to conclude that there are processes which are affected by response repetition alone. Two possible explanations are available. First, repetition may decrease response execution time. This interpretation is contradicted by the consistent findings in Experiments II-IV that READ time is not effected by list length, practice, relative frequency, or sequential effects. Alternatively, the effects could be attributed to the retrieval stage. Retrieval from any response may "activate" an entire area in memory (Collins and

Quillian, 1970; and Meyer and Schavenevelt, 1971), or at least other associations to the same response in this list. The fact that both the response frequency and recency effects are more marked for associated PAs than for non-associated PAs may be indicative of generalization along semantic dimensions. Such "spread of excitation" complicates the model, but points to a necessary component of an adequate representation of recall.

Conclusions

Four experiments have been reported which employed a recall-RT paradigm. A consistent package of results which stands the test of replication emerges from these experiments. First, we will summarize these findings and indicate the questions which require further investigation. Second, we will briefly review the difficulties scanning models have in adequately capturing the processes involved in this task. Third, we will summarize our investigation of a strength model which serves as a useful heuristic in capturing the qualitative results of our experiments. Finally, we will point to the shortcomings of the strength model, and summarize our current theoretical position, specifying prerequisites of models which may provide appropriate explanations of the processes involved in this task, and speculating about the directions one might take in developing and testing an adequate model of retrieval in recall.

Summary of results

The basic observations which have been of interest in the

current work have been the RT-L function, and practice, sequential and frequency effects. In addition, we have attempted to isolate the three processing stages (encoding, retrieval, and execution), and examine the influence of semantic memory in this task. We will briefly summarize our conclusions about each of these points.

Basic findings. The RT-L function is negatively accelerated, although there are some discrepancies in slopes, intercepts, and asymptotes among experiments. Both the slope and intercept of the RT-L function decrease with practice and these effects are largest in early blocks. RT decreases as probe probability increases when manipulated either within or between lists. Also, sequencing of probes is an important variable in this task: RT increases with lag and decreases with consecutive repetitions. Lag does not interact with list length, but the repetition effect is more pronounced in longer lists.

Separation of three stages (encoding, retrieval, and execution). The variables cited above have either no or negligable effects on READ RT. If one is willing to assume that the same encoding and execution processes are involved in the READ and RECALL tasks, all effects can be attributed to the retrieval stage.

Although there do not appear to be any systematic effects attributable to the execution stage (READ task) the results of Experiments III and IV, taken together, demonstrate that frequency and recency of the response produce effects which must be

included in a retrieval model. The exact size and nature of these effects, however, requires further investigation.

Effect of semantic memory (AS variable). A primary interest of the current research has been to investigate the possible use of this recall-RT paradigm to study the effect of the organization of LTM on retrieval. Initially, we have attempted to demonstrate that the associative relationship of the PAs effects RT in a systematic way. While the main and interaction effects involving AS occasionally failed to reach statistical significance, the replicability of these findings across three experiments with different materials is impressive and suggests that the results are reliable. RT is faster to associated PAs, and practice, sequential, and relative frequency effects (whether manipulated between or within lists) are all more marked for non-associates.

The results summarized above have allowed us to draw some conclusions about the adequacies of the various classes of models discussed in the introduction.

Scanning models

Scanning models, which place an emphasis on the role of STM in retrieval, are inadequate to handle our results. The practice and AS effects which were present even when list length was within the capacity of STM indicate that LTM plays a considerable role in retrieval.

Strength models

A simple one-trace strength model which hypothesized RT

to be an inverse function of strength was developed in the context of Experiment I. It was assumed that strength of the S-R association increased when a PA was probed, and decreased when other items were probed. This model makes the strong predictions that: 1) RT should be a negatively accelerated increasing function of list length, and should decrease with increases in probe probability, 2) RT should decrease with practice, 3) the lag function should be negatively accelerated, and 4) RT should decrease with consecutive repetitions of a probe, and this effect should be more marked for longer lists. In addition, with appropriate parameter restrictions the strength model makes the correct prediction that the slope of the RT-L function will decrease with practice. All of these results were observed. Furthermore, when s_{min}, the asymptotic minimum strength an item can decay to, is identified with AS, the model predicts that: 1) RT should be faster to associated than non-associated PAs, 2) the frequency effect (whether manipulated within or between lists) should be more marked for non-associated PAs, and 3) the lag effect should be more marked for non-associated PAs. Again, all of these predictions were confirmed in Experiments II-IV. Thus, the strength model provides a useful integration of a diversity of findings in the present work.

There are several shortcomings of this strength model which point to possible future refinements. First, we found when attempting to fit the data from Experiment I, that we

could not find a suitable combination of parameter estimates which simultaneously predict the prolonged practice and marked sequential effects observed. We argued that this might be a property of a wide class of strength models rather than a problem specific to our formalization. Incorporation of an additional independent trace led to good data fits, but further investigation of duo-process models seems warranted. Second, we found in Experiments III and IV, when probes with two-to-one S-R mappings were included, there were frequency and sequential effects associated with the response which were independent of repetition of the S-R association. Thus, the assumption that the locus of strength changes should be identified with the S-R association seems to be an over-simplification. Finally, there are a number of predictions (e.g., interactions between lag and both list length and practice) which were not verified. Nevertheless, in the absence of consistent opposing trends, we do not consider this to be a major failing of the model.

Future directions

Our inability to develop a one-trace strength model whose parameter values simultaneously yield a prolonged reduction in RT over trials and sequential effects of the magnitude obtained in our data led us to the development of a two-trace model. While this model gave good data fits and reasonable parameter estimates, alternative conceptualizations of the STM process warrant further consideration; although we have modeled the STM process as an additional trace, an STM buffer similar to

the one proposed by Theios, or a single-slot buffer may provide equally good data fits. Formalization of such models and analysis of the differences among them is desirable.

For a model to predict the response effects observed in Experiments III and IV it will be necessary to incorporate retrieval processes independent of the S-R association. There are at least two possibilities. Given that a response has just occurred, it may acrue strength, thus becoming generally more accessible. Alternatively, other S-R bonds involving that response or that entire "area of memory" may be strengthened. Experimentally distinguishing between these possibilities is of interest.

One final point should be considered. In our introduction we stated that our goal was to study the "processes" involved in recall. In our conclusion we have argued that the data support strength models. While such models provide a reasonable mathematical description, one may ask what psychological processes are inherent in them. Perhaps all we can conclude is 1) that the LTM representation of an item is changed in a systematic fashion when items are probed, and 2) that RT is systematically related to the state of the memory representation. These conclusions, however, have lead us to reject scanning models which emphasize the role of STM in retrieval. Furthermore, the strength frame-of-reference leads to the important "process" questions which remain to be answered: What is the nature of the memory representation and the changes

which it encounters, and what are the systematic processes which relate RT to the state of the memory representation?

References

Atkinson, R. C., and Juola, J. F. Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.) Attention and Performance, Vol. IV. New York: Academic Press, 1973.

Atkinson, R. C., and Juola, J. F. Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (Eds.) Contemporary Developments in Mathematical Psychology, in press.

Baddeley, A. D., and Ecob, J. R. Reaction time and short-term memory: A trace strength alternative to the high-speed exhaustive scanning hypothesis. University of California, San Diego Center for Human Information Processing, Report No. 13, 1970.

Baddeley, A. D., and Ecob, J. R. Reaction time and short-term memory: Implications of repetition effects for the high-speed exhaustive scan hypothesis. Quarterly Journal of Experimental Psychology, 1973, 25, 229-240.

Battig, W. F., and Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of Connecticut category norms. Journal of Experimental Psychology Monographs, 1969, 80, (3, Pt. 2), 1-46.

Chandler, J. P. STEPIT: Finds local minima of a smooth function of several parameters. (CPA 312), Behavioral Science, 1969, 14, 81-82.

Clark, H. H. The language-as-fixed effect fallacy: A critique of language statistics in psychological research. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 335-359.

Clifton, C., Jr., and Gutschera, K. D. Hierarchical search of two-digit numbers in a recognition memory task. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 528-541.

Collins, A. M., and Quillian, M. R. Facilitating retrieval from semantic memory: The effect of repeating parts of an inference. Attention and Performance (Vol. III), Acta Psychologica, 1970, 33, 304-314.

Conrad, R. Practice, familiarity and reading rate for words and nonsense syllables. Quarterly Journal of Experimental Psychology, 1962, 14, 71-76.

Corballis, M. C. Serial order in recognition and recall. Journal of Experimental Psychology, 1967, 74, 99-105.

Corballis, M. C., Kirby, J., and Miller, A. Access to elements of a memorized list. Journal of Experimental Psychology, 1972, 94, 185-190.

Corballis, M. C., and Miller, A. Scanning and decision processes in recognition memory. Journal of Experimental Psychology, 1973, 98, 379-386.

Falmagne, J. C. Stocastic models for choice reaction time with application to experimental results. Journal of Mathematical Psychology, 1965, 2, 77-124.

Hawkins, H. L. Mackay, S. L., Holley, S. L. Friedin, B. D., and Cohen, S. L. On the locus of the relative frequency effect

in choice reaction time. Journal of Experimental Psychology, 1973, 101. 90-99.

Homa, D. Organization and long-term memory search. Memory and Cognition, 1973, 1, 369-379.

Howell, W. C. Representation of frequency in memory. Psychological Bulletin, 1973, 80, 44-53.

Kintsch, W. Models for free recall and recognition. In D. A. Norman (Ed) Models of Human Memory. New York: Academic Press, 1970.

Kornblum, S. Sequential effects in choice reaction time: A tutoria review. In S. Kornblum (Ed.) Attention and Performance, Vol. IV, New York: Academic Press, 1973.

Kucera, H., and Francis, W. N. Computational analysis of present-day American English. Providence, R.I.: Brown University Press, 1967.

Meyer, D. E., and Schvanevelt, R. W. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 1971, 90, 227-234.

Myers, J. L. Fundamentals of Experimental Design. (2nd edition), Boston: Allyn and Bacon, 1972.

Naus, M. J., Glucksberg, S., and Ornstein, P. A. Taxonomic word categories and memory search. Cognitive Psychology, 1972, 3, 643-654.

Nickerson, R. S. Binary-classification reaction time: A review of some studies of human information processing capabilities.

Psychonomic Monograph Supplements, 1972, 4 (17, Whole No. 65), 275-318.

Norman, D. A., and Wicklegren, W. A. Strength theory of decision rules and latency in retrieval from short-term memory. Journal of Mathematical Psychology, 1969, 6, 192-208.

Pierce, J. R., and Karlin, J. E. Reading rates and the information rate of a human channel. Bell Systems Technical Journal, 1957, 36, 497-516.

Shiffrin, P. M. Memory search. In D. A. Norman (Ed.) Models of Human Memory. Academic Press, New York, 1970.

Slamecka, N. An examination of trace storage in free recall. Journal of Experimental Psychology, 1968, 76, 504-513.

Smith, E. E. Choice reaction time: An analysis of the major theoretical positions. Psychological Bulletin, 1968, 69, 77-110.

Sternberg, S. High-speed scanning in human memory. Science, 1966, 153, 652-654.

Sternberg, S. The discovery of processing stages: Extensions of Donder's method. Attention and Performance, (Vol. II). Acta Psychologica, 1969(a), 30, 276-315.

Sternberg, S. Memory scanning: Mental processes revealed by reaction-time experiments. American Scientist, 1969(b), 57, 421-457.

Theios, J. Reaction time measurements in the study of memory processes: Theory and data. In G. H. Bower (Ed.) The Psychology of Learning and Motivation (Vol. VII), New York:

Academic Press, 1973.

Theios, J., Smith, P. G., Haviland, S. E., Traupman, J., and Moy, M. E. Memory scanning as a serial self-terminating process. Journal of Experimental Psychology, 1973, 97, 323-336.

Theios, J., and Walter, D. G. Stimulus and response frequency and sequential effects in memory scanning reaction times. University of Wisconsin Mathematical Psychology Program. Report No. 73-1, 1973.

Thorndike, E. L., and Lorge, I. The Teacher's Word Book of 30,000 Words. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon and D. G. Horton (Eds.) Verbal Behavior and General Behavior Theory. Englewood Cliffs, N.J.: Prentice-Hall, 1965.

Tulving, E. Episodic and semantic memory. In E. Tulving and W. Donaldson (Eds.) Organization of Memory. New York: Academic Press, 1972.

Tulving, E., and Madigan, S. A. Memory and verbal learning. Annual Review of Psychology, 1970, 21, 437-484.

Tzeng, O. J. L. A triggering system for scanning processes in long-term memory. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 662-670.

Wicklegren, W. A. Multitrace strength theory. In D. A. Norman (Ed.) Models of Human Memory. New York: Academic Press, 1970.

Wicklegren, W. A. Coding, retrieval, and dynamics of multitrace associative memory. In L. W. Gregg (Ed.) Cognition in Learning and Memory. New York: Wiley, 1972.

Footnotes

[1]This research was supported by an Office of Education grant, OEG-1-71-0109(508) to J. L. Myers and a National Science Foundation grant, NSF GU-4041 to the Department of Psychology. We wish to express our gratitude to David Ellis for his aid in running subjects, and to Nancy A. Myers for her critical comments on an earlier draft of this manuscript.

[2]Mean scores were not computed for individual items. Thus items are essentially treated in our analyses as fixed-effect variables (Clark, 1973); that is, our inferences are formally relevant only to the specific PAs employed. Following Clark's article, we considered reanalyzing our data to treat items as a random-effects variable. However, the cost of reprogramming and loss of statistical power seemed to outweigh the gain in inferential generality. We should note, however, that many of our findings are replicated over several experiments employing different PAs.

[3]For the sake of economy of data presentation, the points in Figure 2 are an unweighted average based on the five lists. A case could readily be made that a weighted average is more appropriate. We chose to use unweighted means because the result is in some sense conservative; weighting lists by the number of observations each contributed to a lag point would have yielded a more marked lag effect. This is because the long RTs

from the longer lists would contribute more to large lags and proportionately less to short lags. By the same reasoning, our test of mean slope referred to in the next paragraph, is also conservative.

[4]This linear mapping permits us to derive closed expressions for a number of statistics of interest (see Appendix I). It may be of interest to note that the model presented in this section is formally equivalent to assuming two states of strength, S_1 and S_2, with associated mean response times, K_1 and K_2, and linear operators on $p_n(S_1)$ and $p_n(S_2)$, the probabilities of being in the states (Falmagne, 1965, presents derivations for a similar model).

[5]We could restate the model in a way which would be formally equivalent but would reduce the number of parameters to be estimated. Assume that each item has an associated retrieval time, t_n, which is "read out" when the item is probed. Upon presentation of an item,

$$t_n = (1 - \alpha_1)t_{n-1}$$

and during the lag

$$t_n = (1 - \alpha_2)t_{n-1} + \alpha_2 t_{max}$$

This model can be shown to be equivalent to our strength model; note that t_{max} equals $k(1-S_{min})$.

[6]We have employed Monte Carlo runs with both logarithmic,

$$T_n = W - k \log S_n$$

and reciprocal,

$$T_n = W + (k/S_n)$$

mappings for various combinations of parameter values. The results are qualitatively consistent with predictions based on the linear mapping, and support our contention that the concommittant description of practice and sequential effects is a general problem for one-trace models.